

Swope, Sheridan

From: Swope, Sheridan
Sent: Wednesday, January 26, 2005 1:34 PM
To: STIC-Biotech/ChemLib
Subject: 10/649,273

For 10/649,273, pls search and interference search:

SID 2, full-length, against the NT and AA data bases
SID 2, residues 148-414, against the NT and AA data bases
SID 2, residues 176-414, against the NT and AA data bases

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
571-272-0943 (voice)
E02B71 Remsen Bld (Office)
E02C70 Remsen Bld (Mailbox)

10/649,273



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 143413

TO: Sheridan Swope
Location: REM-2B71/2C70
Art Unit: 1652
Thursday, February 17, 2005

Case Serial Number: 10/649273

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-A-62
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Swope,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
Remsen 1-A-62
Ext. 22524

From: Swope, Sheridan
Sent: Wednesday, January 26, 2005 1:34 PM
To: STIC-Biotech/ChemLib
Subject: 10/649,273

For 10/649,273, pls search and interference search:

SID 2, full-length, against the NT and AA data bases
SID 2, residues 148-414, against the NT and AA data bases
SID 2, residues 176-414, against the NT and AA data bases

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
571-272-0943 (voice)
E02B71 Remsen Bld (Office)
E02C70 Remsen Bld (Mailbox)

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2- _____
Date Searcher Picked up: _____
Date Completed: _____
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search

NA Sequence: # _____
AA Sequence :# _____
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rnp**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 08:37:51 / Search time 83.25 Seconds
(without alignments)
1923.349 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125
Sequence: 1 MLILTRTAGVFPKPSKRKY.....DISKEVGEASIKVQLKMEI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	5	ABG96478 Novel hum
2	2125	100.0	414	5	ABR05481 Human O-s
3	2125	100.0	414	6	ABJ26654 Human pro
4	2088	98.3	414	5	AAE29234 Human gly
5	2088	98.3	414	6	ABG71161 Novel hum
6	2088	98.3	414	6	ABU09569 Human gly
7	1845	86.8	364	6	ADA54471 Human pro
8	1385	65.2	267	5	ABG96487 Novel hum
9	714.5	33.6	409	4	ABBB69133 Drosophi
10	659.5	31.0	439	3	AAG19286 Arabidops
11	659.5	31.0	444	3	AAAG19285 Arabidops
12	634	29.8	463	3	AAV52216 Arabidops
13	549	25.8	179	5	ABG96489 Novel hum
14	524	24.7	382	6	ABU29234 Protein e
15	524	24.7	421	5	ABG96491 Novel hum
16	502	23.6	348	6	ABU35232 Protein e
17	501	23.6	350	8	ADL065040 M. catarr
18	499.5	23.5	401	7	ABO68626 Pseudom
19	494.5	23.3	343	6	ABU39303 Protein e
20	492.5	23.2	341	4	AAU38205 Pseudom
21	492.5	23.2	341	6	ABU38276 Protein e
22	492.5	23.2	341	7	ADG73342 P aerugin
23	490.5	23.0	341	7	ADG73344 P aerugin
24	489.5	23.0	335	6	ABU27480 Protein e
25	484.5	22.8	337	4	AAU34711 E. coli c

26	484.5	22.8	337	6	ABU47561 Protein e
27	484.5	22.8	337	6	ABU28771 Protein e
28	483.5	22.8	337	6	ABU50237 Protein e
29	482.5	22.7	340	6	ABU40514 Protein e
30	482.5	22.7	357	7	ADP06228 Bacteri
31	481.5	22.7	337	4	AAU38187 Salmonell
32	479.5	22.6	337	3	AAV52204 Escherich
33	475.5	22.4	325	3	AAE26325 Glycoprot
34	475.5	22.4	325	3	AAV52203 Pasteurel
35	475.5	22.4	342	3	AAV52202 Haemophil
36	475.5	22.4	342	4	AAU35450 Haemophil
37	475.5	22.4	342	6	ABU30280 Protein e
38	474.5	22.3	343	6	ABO62704 Klebsiell
39	472.5	22.2	338	6	ABM67812 Phototrab
40	472.5	22.2	363	7	ADC96982 E. faeciu
41	469	22.1	340	6	ABU24130 Protein e
42	468	22.0	338	6	ABU29893 Protein e
43	465.5	21.9	346	3	AAV52208 Bacillus
44	465.5	21.9	354	6	ABU37844 Protein e
45	465	21.9	341	6	ABU40069 Protein e

ALIGNMENTS

RESULT 1	ABG96478	standard; protein; 414 AA.
ID	ABG96478	
XX	ABG96478;	
AC		
XX		
DT	11-DEC-2002	(first entry)
XX		
DE	Novel human metalloprotease Mpl.	
XX		
KM	Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;	Abu47561 Protein e
KM	motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;	Abu28771 Protein e
KM	reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;	Abu50237 Protein e
KM	genital wart; metabolic disorder; premature puberty; Kallman syndrome;	Abu40514 Protein e
KM	Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;	ADP06228 Bacteri
KM	Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;	AAU38187 Salmonell
KM	liver disease; renal disease; immune disorder; rheumatoid arthritis;	AAV52204 Escherich
KM	acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;	AAE26325 Glycoprot
KM	emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;	AAV52203 Pasteurel
KM	neurological disorder.	AAV52202 Haemophil
XX		
OS	Homo sapiens.	AAU35450 Haemophil
XX		
PN	WO200272751-A2.	ABU30280 Protein e
XX		
PD	19-SEP-2002.	ABO62704 Klebsiell
XX		
PF	05-FEB-2002; 2002WO-US003353.	ABM67812 Phototrab
XX		
PR	05-FEB-2001; 2001US-0266518P.	ADC96982 E. faeciu
XX		
PR	10-APR-2001; 2001US-0282814P.	ABU24130 Protein e
XX		
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	ABU29893 Protein e
XX		
PI	Chen J, Feder J, Nelson TC, Duclos F, Kyrstek S;	AAV52208 Bacillus
XX		
DR	WPI; 2002-723329/78.	ABU37844 Protein e
XX		
DR	N-PSDB; ABS76635.	ABU40069 Protein e
XX		
PT	New isolated ameliorating acid encoding MP-1 protein, useful for preventing,	
PT	treating, or ameliorating diseases associated with aberrant	
PT	metalloproteinase activity, e.g. immune, metabolic, inflammatory and	
XX	neurological disorders.	
XX		
PS	Claim 5; Fig 1A-C; 473pp; English.	
XX		
CC	The invention describes an isolated nucleic acid molecule (I) encoding a	
CC	metalloprotease (Mp-1). (I) is useful for preventing, treating, or	

CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloprotease MPI protein

XX Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 2.2e-215; Mismatches 0; Indels 0; Gaps 0;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRKYEFELRSFNFPGTLFLHKIVLGIGTSCDDTAAAVDETSNV 60
 DB 1 MLITKTAGVFFPKSKRKYEFELRSFNFPGTLFLHKIVLGIGTSCDDTAAAVDETSNV 60
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITKPGI 120
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITKPGI 120
 QY 121 ALSGVLSFSLQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180
 DB 121 ALSGVLSFSLQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180
 QY 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGRNHFPIK 240
 DB 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGRNHFPIK 240
 QY 241 PPLHAKNCDFSPFTGLQHTDKIIMKKEKEBEGIEGQILSSAADIAATVQHTMACHLVYR 300
 DB 241 PPLHAKNCDFSPFTGLQHTDKIIMKKEKEBEGIEGQILSSAADIAATVQHTMACHLVYR 300
 QY 301 THRAILPCKORDLIPONNAVIVASGVASNFYIRALETILTNAQTCTLLCPPRLCTDNG 360
 DB 301 THRAILPCKORDLIPONNAVIVASGVASNFYIRALETILTNAQTCTLLCPPRLCTDNG 360
 QY 361 IMIANNGIERLRAGILHIDIEGIRYEPKCPGLVDISKEVGBASIKVPQLKMEI 414
 DB 361 IMIANNGIERLRAGILHIDIEGIRYEPKCPGLVDISKEVGBASIKVPQLKMEI 414

RESULT 2

ABB05481 ID ABB05481 standard; protein; 414 AA.

XX ABB05481;

XX 19-APR-2002 (first entry)

XX Human O-6-ialoglycoproteinase-like protein SEQ ID NO:2.

XX Human; O-6-ialoglycoproteinase-like protein; OSGPLP; enzyme.

XX Homo sapiens.

XX CN1318550-A.

XX 24-OCT-2001.

XX 19-APR-2000; 2000CN-00106834.

XX 19-APR-2000; 2000CN-00106834.

XX

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2002-115090/16.

XX N-PSDB; ABA93268.

PT O-6-ialoglycoproteinase-like protein and encoding polynucleotide, useful
 for diagnosing, preventing and treating related diseases.

XX Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.

The present sequence represents human O-6-ialoglycoproteinase-like protein (OSGPLP). The present invention also describes: (1) the preparation of the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the prevention and/or treatment of related diseases; (4) utilizing the OSGPLP protein in screening its agonist, excitomotor and inhibitor and preparing an antibody against the OSGPLP protein; and (5) the use of the OSGPLP polynucleotide sequences, proteins, agonists, excitomotors, inhibitors and antibodies in treating diseases related to the abnormal OSGPLP gene and in preparing the medicine composite for the treatment

XX Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 2.2e-215; Mismatches 0; Indels 0; Gaps 0;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRKYEFELRSFNFPGTLFLHKIVLGIGTSCDDTAAAVDETSNV 60
 DB 1 MLITKTAGVFFPKSKRKYEFELRSFNFPGTLFLHKIVLGIGTSCDDTAAAVDETSNV 60
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITKPGI 120
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITKPGI 120
 QY 121 ALSGVLSFSLQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180
 DB 121 ALSGVLSFSLQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180
 QY 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGRNHFPIK 240
 DB 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGRNHFPIK 240
 QY 241 PPLHAKNCDFSPFTGLQHTDKIIMKKEKEBEGIEGQILSSAADIAATVQHTMACHLVYR 300
 DB 241 PPLHAKNCDFSPFTGLQHTDKIIMKKEKEBEGIEGQILSSAADIAATVQHTMACHLVYR 300
 QY 301 THRAILPCKORDLIPONNAVIVASGVASNFYIRALETILTNAQTCTLLCPPRLCTDNG 360
 DB 301 THRAILPCKORDLIPONNAVIVASGVASNFYIRALETILTNAQTCTLLCPPRLCTDNG 360
 QY 361 IMIANNGIERLRAGILHIDIEGIRYEPKCPGLVDISKEVGBASIKVPQLKMEI 414
 DB 361 IMIANNGIERLRAGILHIDIEGIRYEPKCPGLVDISKEVGBASIKVPQLKMEI 414

RESULT 3

ABJ26654 ID ABJ26654 standard; protein; 414 AA.

XX ABJ26654;

XX 01-MAY-2003 (first entry)

XX Human protein modification + maintenance molecule protein SEQ ID No. 8.

XX Cyostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;

XX cerebrioprotective; antiparkinsonian; nootropic; antiinflammatory;

XX antitumor; hepatotropic; gynaecological; antibacterial; virucide;

XX protozoacide; antiparasitic; cell proliferative disease; PMOD;

XX protein modification and maintenance molecule; immunogenic fragment;

XX cancer; autoimmune; inflammatory disease; neurological disorder;

KW gastrointestinal; developmental; vesicle trafficking disorder; infection;
 KM protein-protein interaction; drug-target interaction;
 KM gene expression profile; human.
 XX Homo sapiens.
 OS
 XX WC0003000844-A2.
 XX
 PD 03-JAN-2003.
 XX
 PF 18-JUN-2002; 2002MO-US019360.
 XX
 XX 22-JUN-2001; 2001US-0300508P.
 PR 06-JUL-2001; 2001US-0303445P.
 PR 13-JUL-2001; 2001US-0305405P.
 PR 09-AUG-2001; 2001US-0311442P.
 PR 24-AUG-2001; 2001US-0314821P.
 PR 29-AUG-2001; 2001US-0315922P.
 PR 03-MAY-2002; 2002US-0378205P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Gandhi AR, Kabie AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM,
 PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee EA, Yue H;
 PI Forsythe J, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;
 PI Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG;
 PI Malia NK, Mason PM, Gururajan R, Lee S, Beeha SD, Lee SY, Tran UK;
 PI Elliott VS, Luo W, Sprague MW, Tang YT, Lu Y, Zedajadian Y;
 XX
 DR WPI; 2003-184039/18.
 DR N-PSDB; ABT23207.
 XX
 PT New isolated human PMOD polypeptide and polynucleotide, useful for
 PT diagnosing, treating and preventing disease or conditions associated
 PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
 PT infections.
 XX
 PS Claim 63; Page 182-183; 225pp; English.
 XX
 CC The invention relates to an isolated polypeptide comprising: any of 28
 CC sequences of 48-1256 amino acids; a natural amino acid sequence at least
 CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence
 CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
 CC acids, or 97% identical to a sequence of 242 amino acids, all given in
 CC the specification; or a biologically active or immunogenic fragment of
 CC the isolated polypeptide. The polypeptides and polynucleotides are useful
 CC in diagnosing, treating and preventing diseases or conditions associated
 CC with the decreased expression of protein modification and maintenance
 CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
 CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
 CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,
 CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
 CC endometriosis), developmental, vesicle trafficking disorders, and
 CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also
 CC useful in assessing the effects of exogenous compounds on the expression
 CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its
 CC fragments are useful in screening compounds for effectiveness as agonist
 CC or antagonist of the polypeptides, or in altering the expression of the
 CC target polynucleotide and compounds that specifically bind to or modulate
 CC the activity of the polypeptide. The microarray is useful in monitoring
 CC or measuring protein-protein interactions, drug-target interactions, and
 CC gene expression profiles. This sequence represents a human PMOD protein
 CC of the invention
 XX
 SO Sequence 414 AA:
 Query Match 100.0%; Score 2125; DB 6; Length 414;
 Best Local Similarity 100.0%; Pred. No. 2.2e-215;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LGEAHSQTEVALKTGIGVPPAAQOHRENIORIVGEALSASGVSPSDLSAATIKPGI 120
 DB 61 LGEAHSQTEVALKTGIGVPPAAQOHRENIORIVGEALSASGVSPSDLSAATIKPGI 120
 QY 121 ALSIGVGSFSLQIVGQLKKPPIPIHMEAAHLLTRITNKVEFPVLVLLISGHCILALV 180
 DB 121 ALSIGVGSFSLQIVGQLKKPPIPIHMEAAHLLTRITNKVEFPVLVLLISGHCILALV 180
 QY 181 QGVSPDLLGKSLDIPAGDMILKVAARLSLIRPECTSWGGKAIHLLAQGRFFPDIK 240
 DB 181 QGVSPDLLGKSLDIPAGDMILKVAARLSLIRPECTSWGGKAIHLLAQGRFFPDIK 240
 QY 241 PLLHAKNCDPFTGLOHTYDRTKIIMKEKEBEGIKQIISADIAATVQHTACHLVKR 300
 DB 241 PLLHAKNCDPFTGLOHTYDRTKIIMKEKEBEGIKQIISADIAATVQHTACHLVKR 300
 QY 301 TTRAILFCQKQDILLPQNNAVLVAAGGVANFYIRRALILTNATQCTLLCPPRLCTDNG 360
 DB 301 TTRAILFCQKQDILLPQNNAVLVAAGGVANFYIRRALILTNATQCTLLCPPRLCTDNG 360
 QY 361 IMIANNGIRLRAAGLILHDIGIRPERKCPGVDSKVGKASIVPOLKMEI 414
 DB 361 IMIANNGIRLRAAGLILHDIGIRPERKCPGVDSKVGKASIVPOLKMEI 414
 RESULT 4
 AAE29234
 ID AAE29234 standard; protein; 414 AA.
 AC AAE29234;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Human glycoprotease 28472 protein.
 XX
 KW Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
 KW 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arhythmia;
 KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular dysfunction;
 KW hypertension; ischaemic heart disease; obesity; myocardial infarction;
 KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;
 KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;
 KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;
 KW chromosome mapping; tissue typing; gene therapy; neuroprotective;
 KW cytoskeletal; anorectic; cardiac; haemostatic.
 OS Homo sapiens.
 PH
 FH Key
 PH Domain
 FT 1..108 Location/Qualifiers
 FT /note= "Non-transmembrane domain; N-terminal cytoplasmic
 FT domain"
 FT 38..369
 FT /note= "Endopeptidase O-sialoglycoprotein hydrolase
 FT metalloprotease zinc glycoprotease sialoglycoprotease
 FT domain"
 FT 109..132
 FT /note= "Transmembrane domain"
 FT 133..164
 FT /note= "Non-transmembrane domain; non-cytoplasmic loop"
 FT 138..152
 FT /note= "Glycoprotease domain"
 FT 165..189
 FT /note= "Transmembrane domain"
 FT 190..316
 FT /note= "Non-transmembrane domain; cytoplasmic domain"
 FT 317..333
 FT /note= "Transmembrane domain"
 FT 334..414
 FT /note= "Non-transmembrane domain"
 FT 374..414
 FT /note= "Sialoglycoprotease type domain"
 FT
 FT Domain

PN WO200274960-A2.
 XX
 PD 26-SEP-2002.
 XX
 PF 08-NOV-2001; 2001WO-US051427.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
 XX
 DR WPI: 2002-759898/82.
 DR N-PSDB; AAD46856.
 XX
 PT New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
 PT useful for diagnosing and treating cancer, immune, cardiovascular,
 PT hematopoietic, brain, pain, metabolic, liver or platelet disorders, and
 PT in pharmacogenomics.
 XX
 PS Claim 1; Fig 8; 178pp; English.
 XX
 CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
 CC protease or seven transmembrane domain (7TM) receptor family members.
 CC Sequences of the invention are useful in diagnosing and treating cancer
 CC or aberrant cellular proliferation and/or differentiation (e.g. colon or
 CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.
 CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
 CC myocardial infarction, thrombus) including endothelial cell disorders
 CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
 CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
 CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet
 CC disorders. They are also useful in screening assays, predictive medicine
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The
 CC nucleic acids may also be used in chromosome mapping, tissue typing and
 CC forensic biology and as surrogate markers. Sequences of the invention are
 CC also used in gene therapy. The present sequence is human glycoprotease
 CC 28472 protein.
 XX
 SQ Sequence 414 AA:
 Query Match 98.3%; Score 2088; DB 5; Length 414;
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 361 IMIANGIERLPAAGILHDIGIRYEPKCPGVDSISKEVBSATIKVPQAKMEI 414
 DB 361 IMIANGIERLPAAGILHDIGIRYEPKCPGVDSISKEVBSATIKVPQAKMEI 414
 RESULT 5
 ID ABG71161 standard; protein; 414 AA.
 XX
 AC ABG71161;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Novel human glycoprotease 28472.
 XX
 KW Cancer; aberrant cell proliferation; aberrant cell differentiation;
 KW breast cancer; ovarian cancer; prostate cancer; colon cancer;
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;
 KW endothelial disorder; hematopoietic disorder; blood vessel disorder;
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;
 KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;
 KW autoimmune disorder; hypertension; atherosclerosis; heart failure;
 KW myocardial infarction; ischaemic heart disease; Crohn's disease;
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
 KW cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.
 XX
 OS Homo sapiens.
 FH
 FT Key Location/Qualifiers
 FT Domain 138..152
 FT /label= Glycoprotease_domain
 FT
 FT WO200277233-A2.
 FT
 FT 03-OCT-2002.
 FT
 PF 08-NOV-2001; 2001WO-US046724.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
 XX
 DR WPI: 2003-029938/02.
 DR N-PSDB; ABS57020.
 XX
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
 PT hypertension.
 PS
 PS Claim 4; Fig 8A-B; 178pp; English.
 XX
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
 CC sequences that encode a human seven transmembrane domain (7TM). The
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
 CC sequences are useful for diagnosing, preventing or treating a subject
 CC with or at risk of developing a disorder, e.g. cancer or aberrant
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
 CC prostate, colon or lung cancer), immune disorders, heart disorders,
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,
 CC liver disorders or platelet disorders. These disorders include carcinoma,
 CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,

CC hypertension, atherosclerosis, heart failure, myocardial infarction,
 CC ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
 CC cachexia or diabetes. This is the amino acid sequence of the novel human
 CC glycoprotease 28472
 XX

XX Sequence 414 AA;

Query Match 98.3%; Score 2088; DB 6; Length 414;

Best Local Similarity 98.3%; Pred. No. 1.9e-211;

Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLILTKAGVFFPKSKRKVYEFLLRSFNHPGTLFLHKIVLGIEISCDPTAAAVDEGTNV 60
 DB 1 MLILTKAGVFFPKSKRKVYEFLLRSFNHPGTLFLHKIVLGIEISCDPTAAAVDEGTNV 60
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYVEALSASGVSPDLSAIAITTKPGL 120
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYVEALSASGVSPDLSAIAITTKPGL 120
 QY 121 ALSLGVLSFSLQVGVQLKKPFIPIHMEAHALTIRLTNKVFPFPLVLLISGHCILALV 180
 DB 121 ALSLGVLSFSLQVGVQLKKPFIPIHMEAHALTIRLTNKVFPFPLVLLISGHCILALV 180
 QY 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFFPDIX 240
 DB 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFFPDIX 240
 QY 241 PPLHAHAKCDFSPFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 DB 241 PPLHAHAKCDFSPFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 QY 301 THRALIFCKQRDLIPONNAVIVASGVASNTYIRRALIITNAQCTLLCPPRILCTDNG 360
 DB 301 THRALIFCKQRDLIPONNAVIVASGVASNTYIRRALIITNAQCTLLCPPRILCTDNG 360
 QY 361 IMIANGIERLRAGILHDIGIRYEPKCPGVDSKEVGEASIKVPOLKKEI 414
 DB 361 IMIANGIERLRAGILHDIGIRYEPKCPGVDSKEVGEASIKVPOLKKEI 414

RESULT 6

ABU09569 standard; protein; 414 AA.

AC ABU09569;

DT 08-JUL-2003 (first entry)

XX Human glycoprotease encoded by cDNA 28472.

XX Human; enzyme; cancer; aberrant cellular proliferation; differentiation;
 KW immune disorders; heart disorder; brain disorder;
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;
 KW haematopoietic disorder; blood vessel disorder; metabolic disorder;
 KW liver disorder; platelet disorder; glycoprotease.

OS Homo sapiens.

PN US2003009017-A1.

PD 09-JAN-2003.

XX 08-NOV-2001; 2001US-00012140.

XX 08-NOV-2000; 2000US-0246768P.

PR 08-NOV-2000; 2000US-0246772P.

PR 15-NOV-2000; 2000US-0249185P.

XX (LEIBY) LEIBY K R.
 PA (KAPPE) KAPPELLER-LIBERMANN R.
 AC (GLUC) GLUCKSMANN M A.

XX Leiby KR, Kappeller-libermann R, Glucksmann MA;
 PI MPI; 2003-42888/40.
 XX N-PSDB; ACA60887.
 DR

PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic
 PT and liver disorders.

PS Claim 4; Fig 8; 90pp; English.

XX The invention relates to an isolated 38650 (encoding adenosine
 CC deaminase). 28472 (encoding a glycoprotease). 5495 (encoding a 7
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
 CC a sequence which is at least 60% identical to the six nucleic acids or
 CC their open reading frames, fragments of at least 15 nucleotides,
 CC naturally occurring variants, or a DNA insert of the plasmid deposited
 CC with the American Type Culture Collection as Accession No. not defined in
 CC the specification, which encodes the amino acid sequence). Also included
 CC are a host cell containing the nucleic acids (used to produce the
 CC proteins), the encoded proteins, an antibody that selectively binds to
 CC the polypeptide, and identifying a compound that binds to/modulates the
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
 CC methods are useful for diagnosing, treating cancer, aberrant cellular
 CC proliferation and/or differentiation, immune disorders, heart disorders,
 CC cardiovascular disorders including endothelial cell disorders,
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain
 CC and metabolic disorders, liver disorders and platelet disorders (many
 CC examples of these disorders are given in the specification). The present
 CC sequence is the Human glycoprotease encoded by cDNA 28472
 XX

XX Sequence 414 AA;

Query Match 98.3%; Score 2088; DB 6; Length 414;

Best Local Similarity 98.3%; Pred. No. 1.9e-211;

Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLILTKAGVFFPKSKRKVYEFLLRSFNHPGTLFLHKIVLGIEISCDPTAAAVDEGTNV 60
 DB 1 MLILTKAGVFFPKSKRKVYEFLLRSFNHPGTLFLHKIVLGIEISCDPTAAAVDEGTNV 60
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYVEALSASGVSPDLSAIAITTKPGL 120
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYVEALSASGVSPDLSAIAITTKPGL 120
 QY 121 ALSLGVLSFSLQVGVQLKKPFIPIHMEAHALTIRLTNKVFPFPLVLLISGHCILALV 180
 DB 121 ALSLGVLSFSLQVGVQLKKPFIPIHMEAHALTIRLTNKVFPFPLVLLISGHCILALV 180
 QY 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFFPDIX 240
 DB 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFFPDIX 240
 QY 241 PPLHAHAKCDFSPFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 DB 241 PPLHAHAKCDFSPFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 QY 301 THRALIFCKQRDLIPONNAVIVASGVASNTYIRRALIITNAQCTLLCPPRILCTDNG 360
 DB 301 THRALIFCKQRDLIPONNAVIVASGVASNTYIRRALIITNAQCTLLCPPRILCTDNG 360
 QY 361 IMIANGIERLRAGILHDIGIRYEPKCPGVDSKEVGEASIKVPOLKKEI 414
 DB 361 IMIANGIERLRAGILHDIGIRYEPKCPGVDSKEVGEASIKVPOLKKEI 414

RESULT 7

ADA54471 standard; protein; 364 AA.

AC ADA54471;

XX 20-NOV-2003 (first entry)
XX Human protein, SEQ ID 2039.
XX
XX Cytosolic; Anti-inflammatory; Osteopathic; Neuroprotective; Noctropic;
XX Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX inflammatory disease; osteoporosis; neurological disease.
XX Homo sapiens.
XX
XX EPI293569-A2.
XX
XX 19-MAR-2003.
XX
XX 21-MAR-2002; 2002EP-00006586.
XX
XX 14-SEP-2001; 2001JP-00328381.
XX 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Iisogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI; 2003-395539/38.
XX N-PSDB; ADA52832.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
XX and/or membrane proteins, useful for developing medicines for diseases in
XX which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 14; SEQ ID NO 2039; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
XX proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
XX ADA54071). The coding sequences are useful in the gene therapy of
XX diseases caused by abnormalities of the proteins, e.g. cancer,
XX inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 364 AA;
XX
Query Match 86.8%; Score 1845; DB 6; Length 364;
Best Local Similarity 99.2%; Pred. No. 7.9e-186;
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
1 MLILTKAGVFFKSKRVYEFLLSFNHPGTLFLHKLIVLGIETSCDDPTAAVVDGTV 60
1 MLILTKAGVFFKSKRVYEFLLSFNHPGTLFLHKLIVLGIETSCDDPTAAVVDGTV 60
XX
1 LGEAHISQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALISAGVSPDLSAATTIKPGL 120
61 LGEAHISQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALISAGVSPDLSAATTIKPGL 120
XX
1 LGEAHISQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALISAGVSPDLSAATTIKPGL 120
61 LGEAHISQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALISAGVSPDLSAATTIKPGL 120
XX
1 ALSISGVLSTSLQVGVQLKRPPIPHHMEAHALTRLNKVEFPPLVLLISGHCILLV 180
121 ALSISGVLSTSLQVGVQLKRPPIPHHMEAHALTRLNKVEFPPLVLLISGHCILLV 180
XX
1 ALSISGVLSTSLQVGVQLKRPPIPHHMEAHALTRLNKVEFPPLVLLISGHCILLV 180
121 ALSISGVLSTSLQVGVQLKRPPIPHHMEAHALTRLNKVEFPPLVLLISGHCILLV 180
XX
1 QGVSDPFLILGKSLDIPQMDKVARRLSLIKHPECSMSGGKALIEHLAKQGNRHPDIK 240
181 QGVSDPFLILGKSLDIPQMDKVARRLSLIKHPECSMSGGKALIEHLAKQGNRHPDIK 240
XX
1 QGVSDPFLILGKSLDIPQMDKVARRLSLIKHPECSMSGGKALIEHLAKQGNRHPDIK 240
181 QGVSDPFLILGKSLDIPQMDKVARRLSLIKHPECSMSGGKALIEHLAKQGNRHPDIK 240
XX
241 PPIHAAKNCPSFTGLQHVTDKIMKKEKEGIEKQIILSSNADIAATVQHTMAHLVYR 300
241 PPIHAAKNCPSFTGLQHVTDKIMKKEKEGIEKQIILSSNADIAATVQHTMAHLVYR 300
XX
241 PPIHAAKNCPSFTGLQHVTDKIMKKEKEGIEKQIILSSNADIAATVQHTMAHLVYR 300
241 PPIHAAKNCPSFTGLQHVTDKIMKKEKEGIEKQIILSSNADIAATVQHTMAHLVYR 300
XX
301 THRAILFCCKQBDLLPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLCTDNG 360
301 THRAILFCCKQBDLLPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLCTDNG 360
XX
301 THRAILFCCKQBDLLPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLCTDNG 360
301 THRAILFCCKQBDLLPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLCTDNG 360
XX
361 IMIA 364

DB 361 IMIA 364
|||||
RESULT 8
ABG96487
ID ABG96487 standard; protein; 267 AA.
XX
XX ABG96487;
AC
XX
XX 11-DEC-2002 (first entry)
XX
XX Novel human metalloprotease MPI fragment #1.
XX
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
XX motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
XX reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
XX genital wart; metabolic disorder; premature puberty; Kallman syndrome;
XX Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
XX Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
XX liver disease; renal disease; immune disorder; rheumatoid arthritis;
XX acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
XX emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
XX neurological disorder.
XX
XX Homo sapiens.
XX
XX WO200272751-A2.
XX
XX 19-SEP-2002.
XX
XX 05-FEB-2002; 2002MO-US003353.
XX
XX 05-FEB-2001; 2001US-0266518P.
XX 10-APR-2001; 2001US-0282814P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Chen J, Feder J, Nelson TC, Duclos F, Kyrtsek S;
XX
XX WPI; 2002-723329/78.
XX N-PSDB; ABS76639.
XX
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,
XX treating, or ameliorating diseases associated with aberrant
XX metalloproteinase activity, e.g. immune, metabolic, inflammatory and
XX neurological disorders.
XX
XX Claim 5, Page 29; 473pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX metalloprotease (MP-1). (I) is useful for preventing, treating, or
XX ameliorating a medical condition, particularly an immune disorder, an
XX aberrant glutamate transport or motor neuron disorder, such as
XX amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
XX condition. The compositions and methods are also useful for diagnosing,
XX prognosticating, treating, ameliorating and/or treating disorders
XX associated with MP-1 activity, e.g. diabetes, cancer, reproductive
XX disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
XX aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
XX or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
XX Parkinson's disease, Huntington's disease or Tourette syndrome), liver
XX and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
XX or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
XX fibrosis) and vascular, inflammatory and neurological disorders (e.g.
XX Alzheimer's disease or Parkinson's disease). This is the amino acid
XX sequence of a metalloprotease MPI protein
XX
Sequence 267 AA;
XX
Query Match 65.2%; Score 1385; DB 5; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.4e-137;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEAAHATIRLTNNKYPFPPVLVLLISGHCCLLAVGVSDPFLLGKSLDIAPEGMDLKVARR 207
DB 1 MEAAHATIRLTNNKYPFPPVLVLLISGHCCLLAVGVSDPFLLGKSLDIAPEGMDLKVARR 60
QY 208 LSLIKHEPCSTMSGSKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYDKIIMKK 267
DB 61 LSLIKHEPCSTMSGSKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYDKIIMKK 120
QY 268 EKEGIEKGOILSSADIAATVQHTMACHLVKTTRAILFCQKRDLPQNNAVLVASGCV 327
DB 121 EKEGIEKGOILSSADIAATVQHTMACHLVKTTRAILFCQKRDLPQNNAVLVASGCV 180
QY 328 ASNFYIRRAEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGLGLHDIGIRYE 387
DB 181 ASNFYIRRAEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGLGLHDIGIRYE 240
QY 388 PKCPLGVDISKVEGEASIKVPOLKMEI 414
DB 241 PKCPLGVDISKVEGEASIKVPOLKMEI 267

RESULT 9
AB669133 standard; protein; 409 AA.
XX ABB69133;
AC ABB69133;
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 34191.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmacological.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
PA
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI, 2001-656860/75.
XX N-PSDB; ABL13236.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
XX
XX Disclosure; SEQ ID NO 34191; 21pp + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins (AB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 409 AA;
XX
XX Query Match 33.6%; Score 714.5; DB 4; Length 409;

Best Local Similarity 42.0%; Pred. No. 4,4e-66;
Matches 156; Conservative 60; Mismatches 140; Indels 15; Gaps 6;
QY 39 VGIENSQCDTAAAVVDENGVNGEALHSQTEVHLTKGGIVPPAAQOLHRENIQRIYQEA 98
DB 27 VGIETSCDDTGIAVDITGRVYANLYESQOEHTYGGIIPPAQDLHRAIESAYQRC 86
QY 99 LSASGVSPEDLSNATITRIGLALSLGVLSFSLQVLGQKPFIDIHMEAHATIRLT 158
DB 87 MEAAQAKPQVQVLAATTTTRPGLPLSLVGRFARHLARLQKRLVPHHMEAHALQARME 146
QY 159 N-KVEPPLVLLISGHCCLLAVGVSDPFLLGKSLDIAPEGMDLKVARRLSLIKHEPC 216
DB 147 HBEQICVPFLCLLASGHCQLVAVANGPGRLLTLGLQDLDAFGAPKIGRRLHLILPEX 206
QY 217 SMSGGKALIEHLAK-QGNRPHFDIKPPLHAKNCDPSFTGLQHTYDKIIMKKEGIEK 275
DB 207 RLNMGRAIEHAAQLASDPLAYEFPLPLAQORCNPSFAGIKNPSFRAIRARRARTTP 266
QY 276 GQILSSADIAATVQHTMACHLVKTTRAILFC--KQRDLPQNNAVLVASGVAASFYI 333
DB 267 DGVISNYGDFCAGLRSVSRHLMRTQRAIEYCLPRLQDFDTPTLVASGVANNDAL 326
QY 334 RRALBITLNATQCTLLCPPRLCTDNGIMIANGIERLRAGLGLHDIGIRYEKCP 392
DB 327 YANIEHLAAQYGCGRSPRSKRYCSDNGVMIAMHGVQEL-----LQDKASTRYDYD--- 377
QY 393 GVDISKVEGEA 403
DB 378 SIDIOSAGFA 388

RESULT 10
AAG19286 standard; protein; 439 AA.
XX AAG19286;
AC AAG19286;
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21030.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-00301439.
PF
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131445P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132483P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0137882P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.

PR	15-JUN-1999	99US-0144005P
PR	16-JUN-1999	99US-0144085P
PR	16-JUN-1999	99US-0144086P
PR	19-JUN-1999	99US-0144325P
PR	19-JUN-1999	99US-0144331P
PR	19-JUN-1999	99US-0144332P
PR	19-JUN-1999	99US-0144333P
PR	19-JUN-1999	99US-0144343P
PR	19-JUN-1999	99US-0144335P
PR	20-JUN-1999	99US-0144352P
PR	20-JUN-1999	99US-0144632P
PR	20-JUN-1999	99US-0144684P
PR	21-JUN-1999	99US-0144614P
PR	21-JUN-1999	99US-0145088P
PR	21-JUN-1999	99US-0145088P
PR	22-JUN-1999	99US-0145085P
PR	22-JUN-1999	99US-0145087P
PR	22-JUN-1999	99US-0145089P
PR	22-JUN-1999	99US-0145089P
PR	22-JUN-1999	99US-0145118P
PR	22-JUN-1999	99US-0145145P
PR	22-JUN-1999	99US-0145218P
PR	23-JUN-1999	99US-0145224P
PR	23-JUN-1999	99US-0145576P
PR	26-JUN-1999	99US-0145513P
PR	27-JUN-1999	99US-0145518P
PR	27-JUN-1999	99US-0145519P
PR	28-JUN-1999	99US-0145551P
PR	28-JUN-1999	99US-0145551P
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146389P
PR	03-AUG-1999	99US-0147038P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147302P
PR	05-AUG-1999	99US-0147192P
PR	05-AUG-1999	99US-0147260P
PR	06-AUG-1999	99US-0147303P
PR	06-AUG-1999	99US-0147416P
PR	09-AUG-1999	99US-0147933P
PR	09-AUG-1999	99US-0147935P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148319P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148568P
PR	20-AUG-1999	99US-01499292P
PR	20-AUG-1999	99US-01499292P
PR	23-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-01497122P
PR	20-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149864P
PR	23-AUG-1999	99US-0150992P
PR	23-AUG-1999	99US-0150993P
PR	25-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151303P
PR	31-AUG-1999	99US-0151438P
PR	01-SEP-1999	99US-0151930P
PR	07-SEP-1999	99US-0152363P
PR	10-SEP-1999	99US-0153070P
PR	13-SEP-1999	99US-0153758P
PR	15-SEP-1999	99US-0154018P
PR	16-SEP-1999	99US-0154039P
PR	20-SEP-1999	99US-0154779P
PR	22-SEP-1999	99US-0155139P
PR	23-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155569P
PR	28-SEP-1999	99US-0156458P
PR	29-SEP-1999	99US-0156596P
PR	04-OCT-1999	99US-0157113P
PR	05-OCT-1999	99US-0157173P

PR	06-OCT-1999;	99US-0157865P
PR	07-OCT-1999;	99US-0158029P
PR	08-OCT-1999;	99US-0158232P
PR	12-OCT-1999;	99US-0158369P
PR	13-OCT-1999;	99US-0159223P
PR	13-OCT-1999;	99US-0159285P
PR	13-OCT-1999;	99US-0159285P
PR	14-OCT-1999;	99US-0159329P
PR	14-OCT-1999;	99US-0159331P
PR	14-OCT-1999;	99US-0159331P
PR	14-OCT-1999;	99US-0159637P
PR	14-OCT-1999;	99US-0159658P
PR	18-OCT-1999;	99US-0159584P
PR	21-OCT-1999;	99US-0160741P
PR	21-OCT-1999;	99US-0160767P
PR	21-OCT-1999;	99US-0160768P
PR	21-OCT-1999;	99US-0160770P
PR	21-OCT-1999;	99US-0160814P
PR	21-OCT-1999;	99US-0160815P
PR	22-OCT-1999;	99US-0160980P
PR	22-OCT-1999;	99US-0160981P
PR	22-OCT-1999;	99US-0160989P
PR	25-OCT-1999;	99US-0161404P
PR	25-OCT-1999;	99US-0161405P
PR	25-OCT-1999;	99US-0161406P
PR	26-OCT-1999;	99US-0161359P
PR	26-OCT-1999;	99US-0161360P
PR	26-OCT-1999;	99US-0161361P
PR	28-OCT-1999;	99US-0161920P
PR	28-OCT-1999;	99US-0161922P
PR	28-OCT-1999;	99US-0161933P
PR	29-OCT-1999;	99US-0162142P

Query Match 31.0%; Score 659.5; DB 3; Length 444;

[illegible]

RESULT 12	
AAV52216	
ID	AAV52216 standard; protein; 463 AA
XX	
AC	
XX	AAV52216;
DT	09-FEB-2000 (first entry)

XX Arabidopsis thaliana yjld protein homologue.
 DE yjld protein; essential; Gram positive; Gram negative; conserved; motif;
 XX identification; antagonist; antibacterial; antibiotic; broad spectrum;
 KM treatment; infection; resistance; drug target.
 OS Arabidopsis thaliana.
 XX
 XX
 FH Key Location/Qualifiers
 FT Region 86..96 /note= "yjld conserved motif 3"
 FT Region 111..131 /note= "yjld conserved motif 4"
 FT Region 152..198 /note= "yjld conserved motif 2"
 FT Region 208..259 /note= "yjld conserved motif 1"
 FT Region /note= "yjld conserved motif 1"
 XX
 XX MO9954470-A2.
 XX
 XX
 XX 28-OCT-1999.
 XX
 XX 20-APR-1999; 99WO-EP002635.
 XX
 XX 22-APR-1998; 98GB-00008423.
 XX
 XX (GLAX) GLAXO GROUP LTD.
 XX
 XX Arigoni F, Edgerton MD, Lofrer H, Pitsch MC,
 XX
 XX WPI; 2000-013253/01.
 DR
 XX
 PT Novel bacterial polypeptides used to identify broad spectrum antibiotics.
 PS
 PS Claim 1; Fig 1; 55pp; English.
 XX
 CC Proteins AAY52202-Y52217 encompass a novel family of proteins designated
 CC the yjld family, after the name given to the Escherichia coli family
 CC member. These proteins are essential for the survival of both Gram
 CC negative and Gram positive bacteria, although no function has as yet been
 CC ascribed to these proteins. The yjld proteins, fragments of yjld proteins
 CC (for example, fragments encompassing one or more conserved yjld motifs
 CC such as AAY5218-Y52284) and nucleotides encoding them can be used to
 CC identify antagonists and broad spectrum antibacterial compounds. These
 CC antagonists and compounds can be used to treat a wide range of bacterial
 CC infections. New antibiotics are urgently needed, as serious bacterial
 CC infections and antibiotic resistant strains are becoming increasingly
 CC prevalent. The proteins of the invention are essential proteins for
 CC bacterial viability, and represent new targets for antibiotics
 CC
 XX
 XX Sequence 463 AA;
 XX
 XX
 Query Match 29.8%; Score 634; DB 3; Length 463;
 Best Local Similarity 37.9%; Pred. No. 1.8e-57;
 Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;
 QY 38 IVIGIETSCDDTAADVDEGTGVLGEAHSQTEVHLKTGVPPAAQQLHRENIQRIQVE 97
 DB 85 VVIGIETSCDDTAADVDEGTGVLGEAHSQTEVHLKTGVPPAAQQLHRENIQRIQVE 97
 DB 142 ALDPAVNTKRLSLVAVVITIGPGLSLCLFVGVKARVAGNSPSLPIVGVHMEAHALVAL 201
 QY 98 ALSASGVSPSLAIAATIKKGLALSLGVLSFSLQVQLKKPPIPIHHEAAHALTRL 157
 DB 142 ALDPAVNTKRLSLVAVVITIGPGLSLCLFVGVKARVAGNSPSLPIVGVHMEAHALVAL 201
 QY 158 T-NKVEPPLVLLISGHCILALVQVSDPLLIGKSLIAPQMDLKVARRSLIKHPEC 216
 DB 202 VBOELSPFPMALLISGHCILALVQVSDPLLIGKSLIAPQMDLKVARRSLIKHPEC 216
 QY 217 STMSGGAIEPILATCCMMHEDIFELTAAKCUFSFTGLQVHTDKIIMKKEKEGIEKG 276
 DB 259 --RSGGFAVEELALBGDAKSKVKNVPMKTHDCNHSYAGLKTQVLALEAK----- 308

QY 277 QILSSADIAATVQHTMACHLVKTRTRAILFCQKORDLPQNNANVYASGVASNPYIRRA 336
 DB 309 --IRNRADIAASFQRAVAVHLEKCEKRAIDWALE--LEPSIKHWYISGVASNKYVRRL 363
 QY 337 LEILNATQCTLLCPPEPRCTDNGIMIANGTIERAPAGILHDIGIAYE----- 387
 DB 364 LNNIVENKRLKVCPPPSLCTDNGVAVMTGIEHFRVG-----RIDPPPAETPE 413
 QY 388 -----PKCPGLGVDSKEVGEA 403
 DB 414 DYVYDLRPRWPLGEERYAKGRSFA 436
 RESULT 13
 ABG96489
 ID ABG96489 standard; protein; 179 AA.
 XX
 XX ABG96489;
 AC
 XX
 XX 11-DEC-2002 (first entry)
 DT
 XX
 XX Novel human metalloprotease Mpl fragment #3.
 DE
 XX
 XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 KW neurological disorder.
 KW
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200272751-A2.
 PN
 XX
 XX 19-SEP-2002.
 PD
 XX
 XX 05-FEB-2002; 2002WO-US003353.
 PF
 XX
 XX 05-FEB-2001; 2001US-0266518P.
 PR
 XX 10-APR-2001; 2001US-0282814P.
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 PI
 XX
 XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S,
 DR WPI; 2002-723329/78.
 XX
 XX
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
 PT treating, or ameliorating diseases associated with aberrant
 PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
 PT neurological disorders.
 PS
 PS Disclosure; Page 50; 473pp; English.
 PS
 XX
 XX The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.

CC Alzheimer's disease or Parkinson's disease). This is the amino acid
CC sequence of a metalloprotease MPI protein
XX
S0 Sequence 179 AA;
Query Match 25.8%; Score 549; DB 5; Length 179;
Best Local Similarity 66.5%; Pred. No. 3.8e-49;
Matches 119; Conservative 0; Mismatches 0; Indels 60; Gaps 1;
QY 38 IVLGIEISCDPTAAVVDENGVNIGEAHISQTEVHLKTGGIYPPAAQOLHRENIORIVOE 97
DB 1 IVLGIEISCDPTAAVVDENGVNIGEAHISQTEVHLKTGGIYPPAAQOLHRENIORIVOE 60
QY 98 ALSASGVSPDLSAIAITIKPGLALSLGVGSFSLQVLGQKKPPIP----- 144
DB 61 ALSASGVSPDLSAIAITIKPGLALSLGVGSFSLQVLGQKKPPIPCCATTCATCATAT 120
QY 145 -----IHHENAHLLTR 156
DB 121 GGAGGCTCATGCACTTACTATTAGGTGACCATATAAGTAGAATTTCIHHENAHLLTR 179
RESULT 14
ABU22934 ID
ABU22934 standard; protein; 382 AA.
XX
XX ABU22934;
XX
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #8461.
XX
XX Antisense: prokaryotic essential gene; cell proliferation; drug design.
XX
XX Bordetella pertussis.
XX
XX WC200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
XX
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI: 2003-029926/02.
XX
XX N-PSDB; ACA26804.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX
XX for homologous nucleic acids required for cellular proliferation to
XX
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 50858; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX
XX the 6213 antisense sequences given in the specification where expression
XX
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX
XX encoding a polypeptide whose expression is inhibited by the antisense
XX
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX
XX polypeptide or its fragment whose expression is inhibited by the
XX
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX
XX proliferation or the activity of a gene in an operon required for
XX
XX proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibody; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX S0 Sequence 382 AA;
Query Match 24.7%; Score 524; DB 6; Length 382;
Best Local Similarity 37.3%; Pred. No. 5.7e-46;
Matches 134; Conservative 63; Mismatches 136; Indels 26; Gaps 10;
QY 29 HFG--TLFLHK--IVLGIEISCDPTAAVVDENGVNIGEAHISQTEVHLKTGGIYPPAA 83
DB 24 HPGPRTLNVNSAPMIITLGFESSCDEGVAAVCTERGLLAHLALTQTAMHQYGVVBELA 83
QY 84 QOLHRENIORIVOEALNSAGVSPDLSAIAITIKPGLALSLGVGSFSLQVLGQKKPPI 143
DB 84 SRDHLRRVPLTRQVLAENGTLTADVGAVAYTAGPGLAQLLVGASVAQDLANSRALPAL 143
QY 144 PIHHEAHALITRLTN-KVEPFLVLLISGCHLALVGVSPDLSLIGKSLDIPADMDL 202
DB 144 GIHHEGHLISPLAEPREPFVALVSGHTQLMLVGVGRYELGRTLDAAAEAPD 203
QY 203 KVARRLSLIKPECSGTSMSGKAIIEHLAKQGNRPHDIKPELHAKKNDPSFTGLQ-HVTD 261
DB 204 KSAKLMGL-GYP-----GGPALRLAEGQDSRSLPRPMLHSGLDPSFSLKTAVALT 256
QY 262 KIKMKKEKBEIEKQIISLADIAATVGHMACHVKXTHRALILFCRQDLIPQNNAYL 321
DB 257 RV--RAATFDGGELEB--ODRADLAAATQAAVVEVLAARAIKRL---KQTGL-----RRL 304
QY 322 VASGGVANSFYIRALIEILTNATCCTLCPPRLCTDNGIMTAMNGIERLRAGLGIHD 380
DB 305 VVAGGVANALIRAHLARALKFLRAEAYFPPLSLCTDNGAMLAFAAEKVKAGLADIRE 363
RESULT 15
ABG96491 ID
ABG96491 standard; protein; 421 AA.
XX
XX ABG96491;
XX
XX 11-DEC-2002 (first entry)
XX
XX Novel human metalloprotease associated protein #2.
XX
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
XX
XX motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
XX
XX reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
XX
XX genital wart; metabolic disorder; premature puberty; Kallman syndrome;
XX
XX Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
XX
XX Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
XX
XX liver disease; renal disease; immune disorder; rheumatoid arthritis;
XX
XX acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
XX
XX emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
XX
XX neurological disorder.
XX

Db 368 IYAHARSIDICTDASSI-----IDTPRKL 392
 Search completed: February 16, 2005, 13:05:54
 Job time : 91.25 secs

OS Caenorhabditis elegans.
 PN WO200272751-A2.
 XX
 PD 19-SEP-2002.
 PF 05-FEB-2002; 2002WO-US003353.
 XX
 PR 05-FEB-2001; 2001US-0266518P.
 PR 10-APR-2001; 2001US-0282814P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Chen J, Feder J, Nelson TC, Duclos F, Kryatek S;
 PI WPI; 2002-723329/78.
 DR
 XX
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
 PT treating, or ameliorating diseases associated with aberrant
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
 PT neurological disorders.
 XX
 PS Disclosure; Page 465-466; 473pp; English.
 XX
 CC The invention describes an isolated nucleic acid molecule (1) encoding a
 CC metalloproteinase (MP-1). (1) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloproteinase MP1 associated protein
 XX
 SQ Sequence 421 AA;

Query Match 24.7%; Score 524; DB 5; Length 421;
 Best Local Similarity 33.4%; Pred. No. 6.7e-46;
 Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;

QY 39 VLGIETSCDDPTAAAVVDCTGVLEAITHSOTREVLKKTGIVPPAAQQLHRENIQIVGEA 98
 DB 25 VLGIETSCDDPTAAVAVIVKEKEILSSERYTERAIORQGGINPSVCALQHRENLPRILIKC 84
 QY 99 LSAGVSPDLSAATYTIKQGLALSLGVLSFSLQVLQKPFPIHHMEAHALTIRLT 158
 DB 85 LNDAGTSPKLDNAVAVTTPGLVALKEGISAIGPAKKHRLPLIPVHHMGAHALSLILV 144
 QY 159 -NKYEPFVLVLLISGGHCLALVGVSDPFLLGKSLDIAPGMDLDKVARSLINHPES 217
 DB 145 DDEVRFPFSAVLISGGHALISVADVEKFKLYGQSVSGSPCECIDKVARQLDL-GSEPD 203
 QY 218 TMSGGKALIEHLAQGN--RHPDIKPRLHHAKNCDPSFTGLQHVTDKIIMKEKEBIE 274
 DB 204 GIHVGAANVELILASASADGHLRYPIPLPNPKANNPFQIKGSYLNLLERLKNSETSID 263
 QY 275 KGOILSSADIAATVOTMACHLVKTRTHAILFCORDLIPONNAVLVASGGVASFYIR 334
 DB 264 -----IPDFCASLQNTVAHISSKLHIFPESLSQEKLPKQ--LVIGGVANQYIF 313
 QY 335 RALEILTNATQCTLLCPPLCTDNGIMIANNGIERLAGLILIDIGIRYRP----- 388
 DB 314 GAISKLSAANVTITIKVLLSLCTDNAMEMIAVSGL-----LMLVRSRAIMWRPNDIDPT 367
 QY 389 -----KCPPLGVNDISKVEGASIKVPQLKM 412

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:43:41 ; Search time 24.75 Seconds
(without alignments)
1248.675 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125
Sequence: 1 MLILTKTAGVFFPKSKRKRVY.....DISKEVGEASIKVPOLKMEI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pdp:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pdp:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pdp:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pdp:*
5: /cgn2_6/ptodata/1/iaa/PTUS.COMB.pdp:*
6: /cgn2_6/ptodata/1/iaa/backfilest.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	US-10-067-443-2	Sequence 2, Appli
2	2090.5	98.4	439	US-10-067-443-19	Sequence 19, Appl
3	1385	65.2	267	US-10-067-443-32	Sequence 22, Appl
4	634	29.8	463	US-10-067-443-3	Sequence 3, Appli
5	549	25.8	179	US-10-067-443-25	Sequence 25, Appl
6	524	24.7	421	US-10-067-443-4	Sequence 4, Appl
7	524	24.7	421	US-10-067-443-28	Sequence 28, Appl
8	501	23.6	350	US-09-540-236-2726	Sequence 2726, Ap
9	499.5	23.5	401	US-08-252-991A-17372	Sequence 17372, A
10	488	23.0	342	US-08-087-797-3	Sequence 3, Appli
11	482.5	22.7	357	US-09-543-681A-6513	Sequence 6513, Ap
12	475.5	22.4	325	US-08-087-797-2	Sequence 2, Appli
13	474.5	22.3	343	US-09-489-039A-9221	Sequence 9221, Ap
14	472	22.2	363	US-09-107-532A-6609	Sequence 6609, Ap
15	461	21.7	336	US-08-987-121A-4	Sequence 4, Appli
16	457	21.5	327	US-10-067-443-5	Sequence 5, Appli
17	457	21.5	335	US-08-961-083-52	Sequence 52, Appl
18	457	21.5	335	US-09-536-784-52	Sequence 52, Appl
19	455	21.4	336	US-09-066-512-2	Sequence 2, Appli
20	454	21.4	366	US-09-114-000C-4956	Sequence 4956, Ap
21	453	21.3	336	US-09-583-110-4857	Sequence 4857, Ap
22	448	21.1	336	US-09-107-433-4221	Sequence 4221, Ap
23	443	20.8	368	US-09-134-001C-3909	Sequence 3909, Ap
24	433	20.1	341	US-03-145-624-2	Sequence 2, Appli
25	406.5	19.1	344	US-09-198-452A-213	Sequence 213, App
26	406.5	19.1	360	US-09-438-185A-196	Sequence 196, App
27	401.5	18.9	344	US-09-602-777A-148	Sequence 148, App

28	362	17.0	292	4	US-09-724-623-81	Sequence 81, Appl
29	359	16.9	214	4	US-09-328-352-4609	Sequence 4609, Ap
30	338.5	15.9	340	4	US-10-067-443-6	Sequence 6, Appli
31	319	15.0	273	4	US-09-710-279-728	Sequence 728, App
32	204	9.6	187	4	US-09-248-796A-19489	Sequence 19489, A
33	181.5	8.5	143	4	US-09-328-352-4387	Sequence 4387, Ap
34	109	5.1	515	4	US-09-252-991A-33046	Sequence 33046, A
35	109	4.7	225	4	US-09-602-777A-144	Sequence 144, App
36	97.5	4.6	580	4	US-09-489-039A-12013	Sequence 12013, A
37	97	4.6	601	3	US-09-155-036-5	Sequence 5, Appli
38	97	4.6	601	4	US-09-866-307-5	Sequence 5, Appli
39	97	4.6	614	3	US-09-155-036-6	Sequence 6, Appli
40	97	4.6	614	4	US-09-866-307-6	Sequence 6, Appli
41	96	4.5	612	4	US-09-902-540-13545	Sequence 13545, A
42	94.5	4.4	480	4	US-09-583-110-5050	Sequence 5050, Ap
43	94.5	4.4	481	4	US-09-107-433-3197	Sequence 3197, Ap
44	93.5	4.4	294	4	US-09-107-532A-6933	Sequence 6933, Ap
45	93.5	4.4	579	4	US-09-543-681A-5019	Sequence 5019, Ap

ALIGNMENTS

RESULT 1									
US-10-067-443-2									
; Sequence 2, Application US/10067443									
; Patent No. 6642041									
; GENERAL INFORMATION:									
; APPLICANT: Bristol-Myers Squibb Company									
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN									
; FILE REFERENCE: D0073 NP									
; CURRENT APPLICATION NUMBER: US/10/067,443									
; CURRENT FILING DATE: 2002-02-05									
; PRIOR APPLICATION NUMBER: US 60/266,518									
; PRIOR FILING DATE: 2001-02-05									
; PRIOR APPLICATION NUMBER: US 60/282,814									
; PRIOR FILING DATE: 2001-04-10									
; NUMBER OF SEQ ID NOS: 71									
; SOFTWARE: Patent in version 3.0									
; SEQ ID NO 2									
; LENGTH: 414									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-067-443-2									
Query Match 100.0%; Score 2125; DB 4; Length 414;									
Best Local Similarity 100.0%; Pred. No. 1,4e-234;									
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MLILTKTAGVFFPKSKRKRVYEFIRSFNPHGTLPLHKIVLGIETSCDDPAAVVDETGAV	60						
DB	1	MLILTKTAGVFFPKSKRKRVYEFIRSFNPHGTLPLHKIVLGIETSCDDPAAVVDETGAV	60						
QY	61	IGBAHSQTEVHLKKGIVIPPAOOLHRENIRIVGBASASGVSSDSALATTKRGL	120						
DB	61	IGBAHSQTEVHLKKGIVIPPAOOLHRENIRIVGBASASGVSSDSALATTKRGL	120						
QY	121	ALSLGVGSFSLQVLGOLKKEPTPIHMEAHALLTRITNKVEPPVLVLSGGHLLALV	180						
DB	121	ALSLGVGSFSLQVLGOLKKEPTPIHMEAHALLTRITNKVEPPVLVLSGGHLLALV	180						
QY	181	QGVSDPLLGKSLDIAFGDMLDKVARRSLIKHPECSTWGGKAIIEHLAKOGRFFFDIK	240						
DB	181	QGVSDPLLGKSLDIAFGDMLDKVARRSLIKHPECSTWGGKAIIEHLAKOGRFFFDIK	240						
QY	241	PLIHAACNDPSFTGLQHTVDKIIMKKKEBGIKKGQIISSAADIAATVQHTMACHLVR	300						
DB	241	PLIHAACNDPSFTGLQHTVDKIIMKKKEBGIKKGQIISSAADIAATVQHTMACHLVR	300						
QY	301	THRALTPCKQDRLPONNAVIVASGVASNFYIRRLLETITNAVTOCTLLCPPRLCTDNG	360						
DB	301	THRALTPCKQDRLPONNAVIVASGVASNFYIRRLLETITNAVTOCTLLCPPRLCTDNG	360						

QY 361 IMIANGIERLRAGILIHDIIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 414
 Db 361 IMIANGIERLRAGILIHDIIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 2

US-10-067-443-19
 ; Sequence 19, Application US/10067443

; Patent No. 6642041

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED

; FILE REFERENCE: D0073 NP

; CURRENT APPLICATION NUMBER: US/10/067,443

; CURRENT FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: US 60/266,518

; PRIOR FILING DATE: 2001-02-05

; PRIOR APPLICATION NUMBER: US 60/282,814

; PRIOR FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 71

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 19

; LENGTH: 439

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-067-443-19

Query Match 98.4%; Score 2090.5; DB 4; Length 439;
 Best Local Similarity 93.8%; Pred. No. 1.4e-230;

Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVPFKSKRKYVEFLRSFNHFGTLFLHKTIVLGITISCDTAAAVDETGNV 60
 Db 1 MLITKTAGVPFKSKRKYVEFLRSFNHFGTLFLHKTIVLGITISCDTAAAVDETGNV 60
 QY 61 LGEAHSOTVHLKTGIVPPAAOQHRNTOIVQELASGVSPDLSALATTIKFGL 120
 Db 61 LGEAHSOTVHLKTGIVPPAAOQHRNTOIVQELASGVSPDLSALATTIKFGL 120
 QY 121 ALSLGVLSFSLQVGLQKPFPIHMEHAHALTIRLNKVEPPLVLLISGHCILALV 180
 Db 121 ALSLGVLSFSLQVGLQKPFPIHMEHAHALTIRLNKVEPPLVLLISGHCILALV 180
 QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPBCSTWSGKALEHLAKQNRHFPIK 240
 Db 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPBCSTWSGKALEHLAKQNRHFPIK 240
 QY 241 PPLHANKCPFSFTGLQHVTDKTIIMKEKEKEGI-----EK 275
 Db 241 PPLHANKCPFSFTGLQHVTDKTIIMKEKEKEGI-----EK 275
 QY 276 GQILSSADIAIATVQHTMACLVKTRTHRAILFCQKRDILLPNNNAVIVASGVASNFYIR 335
 Db 301 GQILSSADIAIATVQHTMACLVKTRTHRAILFCQKRDILLPNNNAVIVASGVASNFYIR 360
 QY 336 ALBILTNATQCTLLCPRPRLCTDNGIMIAMNGIERLRAGILIHDIIEGIRYBPCPLGVD 395
 Db 361 ALBILTNATQCTLLCPRPRLCTDNGIMIAMNGIERLRAGILIHDIIEGIRYBPCPLGVD 420
 QY 396 ISKEVGEASIKVPOLKMEI 414
 Db 421 ISKEVGEASIKVPOLKMEI 439

RESULT 3

US-10-067-443-22

; Sequence 22, Application US/10067443

; Patent No. 6642041

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED

; TITLE OF INVENTION: SPINAL CORD, MP-1
 ; FILE REFERENCE: D0073 NP
 ; CURRENT APPLICATION NUMBER: US/10/067,443
 ; CURRENT FILING DATE: 2002-02-05
 ; CURRENT APPLICATION NUMBER: US 60/266,518
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: US 60/282,814
 ; PRIOR FILING DATE: 2001-04-10
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 267
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-067-443-22

Query Match 65.2%; Score 1385; DB 4; Length 267;
 Best Local Similarity 100.0%; Pred. No. 5.1e-150;

Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEAHALTRLTNKVEPPLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207
 Db 1 MEAHALTRLTNKVEPPLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60
 QY 208 ISLKHPECSTWSGKALEHLAKQNRHFPIKPLHAKNCDPFTGLQHVTDKTIIMK 267
 Db 61 ISLKHPECSTWSGKALEHLAKQNRHFPIKPLHAKNCDPFTGLQHVTDKTIIMK 120
 QY 268 EKEGIEKGQILSSADIAIATVQHTMACLVKTRTHRAILFCQKRDILLPNNNAVIVASGV 327
 Db 121 EKEGIEKGQILSSADIAIATVQHTMACLVKTRTHRAILFCQKRDILLPNNNAVIVASGV 180
 QY 328 ASNFYIRRALETITNAQCTLLCPRPRLCTDNGIMIAMNGIERLRAGILIHDIIEGIRY 387
 Db 181 ASNFYIRRALETITNAQCTLLCPRPRLCTDNGIMIAMNGIERLRAGILIHDIIEGIRY 240
 QY 388 PKCPLGVDISKVEGASIKVPOLKMEI 414
 Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 4

US-10-067-443-3

; Sequence 3, Application US/10067443

; Patent No. 6642041

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED

; FILE REFERENCE: D0073 NP

; CURRENT APPLICATION NUMBER: US/10/067,443

; CURRENT FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: US 60/266,518

; PRIOR FILING DATE: 2001-02-05

; PRIOR APPLICATION NUMBER: US 60/282,814

; NUMBER OF SEQ ID NOS: 71

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 3

; LENGTH: 463

; TYPR: PRT

; ORGANISM: Arabidopsis thaliana

US-10-067-443-3

Query Match 29.8%; Score 634; DB 4; Length 463;
 Best Local Similarity 37.9%; Pred. No. 1.8e-63;

Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

QY 38 IVLGITSCDDTAAAVDETGNVLGEAHSQTEVHLKTGIVPPAAOQHRNTOIVQELASGV 97
 Db 85 IVLGITSCDDTAAAVDETGNVLGEAHSQTEVHLKTGIVPPAAOQHRNTOIVQELASGV 141
 QY 98 ALSAGVSPDLSALATTIKFGLALSIGVLSPSLDVQGLKPFPIHMEHAHALTR 157

Db	142	ALDKANLEKODLSAVAVTIGPGLSTLCRARVGRARBARVGFSLPIGVHHMEHNAVLARL	201
Qy	158	T-NKVEFPFVLLIGSGHCLTLVQGVSPFLIGKSLDIAFGMDLKVARLSITKPEC	216
Db	202	VEQELSPFPFMAILLISGHNLLVLAHQCGYTQGLGTVDVAIGEAFDKTAKMLGIDMH---	256
Qy	217	STMSSGKKILEHLAKONRPFHDIKPRLHAKANDSPFGLOHTYDRIKMKKEBEGIEKG	276
Db	259	--RSGCPRAVEELALBEDDASVYKFNPMKHKDQCNFSYAGLKTQVRLAITEKE-----	306
Qy	277	QILSSAADIATVQHTMACHLVTKRTHRAILLFCQRDILLPQNNALVLAASGVASNFYIRRA	336
Db	309	--IRNRADIAASFQRAVALHLEBKERCALIDMALE---LPSISGHWISGGVASNKYVRLR	365
Qy	337	LEILTNALQCTLLCEPPRLCTDNGIMIANGLIRLRLAAGLGLHDIGIRYE-----	387
Db	364	LNNIIVENKQLKVLCEPPSLCTDNGVWVAWMTGLEHFRVGG-----RYDPPPEATEPE	413
Qy	388	-----PKCPGLGVDSKEVEGA 403	
Db	414	DYVDLRPRMPLGEVEYAKGRSBA 436	

```

RESULT 5
US-10-067-443-25
/ Sequence 25 Application US/10067443
/ Patent No. 6642041
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
/ TITLE OF INVENTION: SPINAL CORD, NP-1
/ FILE REFERENCE: D0073 NP
/ CURRENT APPLICATION NUMBER: US/10/067,443
/ CURRENT FILING DATE: 2002-02-05
/ PRIOR APPLICATION NUMBER: US 60/266,518
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: US 60/282,814
/ PRIOR FILING DATE: 2001-04-10
/ NUMBER OF SEQ ID NOS: 71
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 25
/ LENGTH: 179
/ TYPE: PRT
/ ORGANISM: homo sapiens
/ US-10-067-443-25

```

RESULT 6
 US-10-067-443-4
 ; Sequence 4, Application US/10067443
 ; Patent No. 6642041
 ; GENERAL INFORMATION:
 ; APPLICANT: Bristol-Myers Squibb Company;
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
 ; TITLE OF INVENTION: SPINAL CORD, NP-1
 ; FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443
 CURRENT FILING DATE: 2002-02-05
 PRIOR APPLICATION NUMBER: US 60/266,518
 PRIOR FILING DATE: 2001-02-05
 PRIOR APPLICATION NUMBER: US 60/282,814
 PRIOR FILING DATE: 2001-04-10
 NUMBER OF SEQ ID NOS: 71
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 4
 LENGTH: 421
 TYPE: PRT
 ORGANISM: Caenorhabditis elegans
 US-10-067-443-4

Query Match 24.7%; Score 524; DB 4; Length 421;
 Best Local Similarity 33.4%; Pred. No. 6.3e-51;
 Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;

QY VLGIEITSCDDTAAAVDEETGNVLSBAIHSGTEVHLKTGGIVPPAAQQLHRENIQRIYQEA 98
 DB |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 84
 25 VLGIEITSCDDTAAVALYNEKREILSSERYTERAIQROGGGINPSVCAQLQHRENLRLIEKC 84
 QY LSASGVSPSDLSAATTTKPGIALSLGVGLSFLQVLGQLKKPPIPIHMEAAHLYTRLT 158
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 144
 85 LINDACTSPDLDAVVAVTVTPGLVIALKEGISAIGFAKRRLLPIPIHMEAAHLYTRLT 144
 QY 159 -NKVEPFLVLLISGGHCLAVQGVSDPFLLSKSLDIAPGDMIDKVARRLSLIKHPECS 217
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 203
 145 DDSVAFPPSAVLLSGGHALISVAEDVEKRLYQSVSGSGBCIDKVARQLGDL-GSEPD 203
 QY 218 TMSGKALIEHLAKQGN--RFHFDIKPPLHAKNCDSPFTGLQHTVDKIIIMKKEKEGIE 274
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 263
 204 GIHVGAIVEIILSRASADGHLRYPIPLPNVPKANNMFDQKGSYTLNLERLRKNSETSID 263
 QY 275 KQILISSAADIAATYQHTMACHLVKTRHAILFCRQDILLPQNNAAVLVAAGVANSFTIR 334
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 313
 264 -----IPDFCASIQNTVARRHISKLMHFFESLSSEQEKLPKO--LVIGGVANQYIF 313
 QY 335 RALBITLVATQCTLLCPPEPLCTDNGIMIAMNGIERLRAGILHDIEGIRYEP----- 388
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 367
 314 GAISTLSAHHVTTIKVLLSLCTDNAMEIMAYSGL-----LMLVNRSEALIMRNDIPDT 367
 QY 389 ----KCPPLGVDISKEVGEASIKVPOLKM 412
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 392
 368 IYAHARSDIGTDASSEI-----IDTPERKL 392

RESULT 7
 US-10-067-443-28
 Sequence 28, Application US/10067443
 Patent No. 6642041
 GENERAL INFORMATION:
 APPLICANT: Bristol-Myers Squibb Company
 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
 TITLE OF INVENTION: SPINAL CORD, MP-1
 FILE REFERENCE: D0073 NP
 CURRENT APPLICATION NUMBER: US/10/067,443
 CURRENT FILING DATE: 2002-02-05
 PRIOR APPLICATION NUMBER: US 60/266,518
 PRIOR FILING DATE: 2001-02-05
 PRIOR APPLICATION NUMBER: US 60/282,814
 PRIOR FILING DATE: 2001-04-10
 NUMBER OF SEQ ID NOS: 71
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 28
 LENGTH: 421
 TYPE: PRT
 ORGANISM: Caenorhabditis elegans
 US-10-067-443-28

```

; ORGANISM: Caenorhabditis elegans
US-10-057-443-29

Query Match      24.7%; Score 524; DB 4; Length 421;
Best local Similarity 33.4%; Pred. No. 6.3e-51;
Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8

```

```

QY 39 VLGJETS CDDPRAA VBE TENV LGAHIS IORE VHLKTG GIVP PPAQO L HRENI O RY OEA 98
Db 25 VLGJETS CDDPRAA VBE TENV LGAHIS IORE VHLKTG GIVP PPAQO L HRENI O RY OEA 98
QY 99 LSA GVS PPSDLSA IATTIKPGALSLAGVLSQLVGLVKKPEPIPHHMAHALTTIRLT 158
Db 85 LNDGATSPKDLDAVA VTTVPGVIALKEGISALIPAKKHRLPLIPVHHMAHALSTILV 144
QY 159 -NKVEP PVLVILJGGHGLALVGVSPFLLLKGS LDIAPDMDKVARRLSLKHPECS 217
Db 145 DDVSRAFPFSAV LLSGGHALISVADVEKFXLYGQSVSSPESCDIKVAROLGDI-GSEFD 203
QY 218 TMSGKXIEHLAKGN--REFPDIKPEPLHAKXCDPESFTGLJOHTDKIIMKKEKEBIE 274
Db 204 GIHVGA VEILLASRA SADGHLRYPIFLPNVKANMPQIQGSTYLNLERLRKSSISID 263
QY 275 KGOILSSAADIAATVOHTMACHLVKRTTRAILPFCKORDLPONNAVIVASGVASNFYIR 334
Db 264 -----IPDFCASIQNTVA RHISKLHIFPESISEGKPKQ--LYIGGVANQYTF 313
QY 335 RALEILMAVOC TLLCPPEPLCTDNGIMIANNGIERLPAAGIGLHIDIEGIRYEP----- 388
Db 314 GAISKLSAAHVVTITIKVLLSLCTDNAMEIMAYSGL-----LMLVNRSEALIMRPNDIPDT 367
QY 389 -----KCELVGDISKEVGEASIRKVPOLCK 412
Db 368 IYAHARSDIGTDASSEI-----IDTPRRKJ 392

```

```

RESULT 8
US-09-540-236-2726
; Sequence 2726, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Bretton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATALYTICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709/2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2726
; LENGTH: 350
; TYPE: PR1
; ORGANISM: M.cattarrhalis
US-09-540-236-2726

23.6%; Score 501; DB 4; Length 350;
Query Match Best Local Similarity 34.5%; Pred. No.2e-48;
Matches 126; Conservative 56; Mismatches 139; Indels 44; Gaps 8;

QY 39 VLGIETSCDDTAAAVDET---GNVLGEAIIHSQTEVHLTKGTGIYPPAAQLHRENTORI 94
DB 6 VLGIETSCDEDEGLAIYDSTVMNRRGGVLSQVYLSQINLHATGCGVPPELASRDHRIKRVPL 65
QY 95 VOEALASGVSPPSLASAIATIKPGLASLGVGSFSQLQVLGQIKKPEIPRIHHNEAAHAT 154
DB 66 FNEMLDDQNNITKSDIDAVALTKGPGELIGALMTGALFGRTLLYGLGVPAVGVHMEGHLLA 125
QY 155 IRLT-NKVEPPFLVTLISGGHCLLALVQGVSDPLLKSLDIAFGMDLDRVARLSLILK 212
DB 126 PLASDDPPSPFPFCLLVSGGHITMLVRADGVGVQIIGESLDDAVGECFDFTAKMLTK-P 184
QY 213 HPECSNMSGKALIEHLAAQGNRFHFDIRPLHAAKNCDFSTGLQHTDTKIMKKKEEG 272
DB 185 YP-----GGNIEKDLAQNPNHAYELRPMQH-KGLDFSPSGMKTALHNDIKQTPNQS 237
QY 273 IEKGQIISSADIIATVQHTMACHLVKRTTHALLFCQKRDLLPQNNATLVASGGVASFY 332
DB 238 DP-----ATRADIAASFAYAVDTLVKKCTYALQMTGIRQ-----LVVAGGSANQT 284
QY 333 IRRALLETITNATQCTLLCPPLRCTDNGIMTANNGIERLRAG-----L 375

```

```

Db      | 285 LRFPLTFETLTAQIDASVYVAATPELTCTDNGAMTAVAGFCRLSRGSGDDLAVACIFRMDWTML 344
       |||
Oy      | 376 GILHD 380
       || : 
Db      | 345 GIEYD 349

RESULT 9
US-09-252-991A-17372
; Sequence 17372, Application US/09252991A
; Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIORITY FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17372
LENGTH: 401
TYPE: PRF
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17372

Query Match          23.5%; Score 499.5; DB 4; Length 401;
Beet Local Similarity 34.4%; Pred. No. 3.ee-48;
Matches 133; Conservative 67; Mismatches 158; Indels 29; Gaps 10;

Oy      | 13 KPSKRYEFLRSFNPHFTLFHKI-----VLGIETSCDDTPAAAVVDGTGVNIGEAIH 66
        |||||:::||:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Db      | 31 KASDRKTLYWRKAADYHARLARLRICQALRLXRMRVLGLETSCDEDTGVNLYDBERGLADALF 90
        ::|||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Oy      | 67 SOTEVHLTKTGIVPPAQOLAHRENIORIYOELASASGSVPDSLSAIATTIKGLIALSLGY 126
        ||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Db      | 91 SQIDLHRVGWGPVELASRDHVKKMPLINQLVDESCTPADIDAITYTAGPGLVGALLV 150
        ||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Oy      | 127 GLSSTLOVGOKKPPIPIHHMEHALTTITNK-VFPPLVVLLISGHCLLAIVQSVD 185
        ||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Db      | 151 GASCAQMMAAWGWPVAGVHHMHGHTLAPMLERPFRPFVALLVSGHQTOLVRVNDIGR 210
        ||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Oy      | 186 FLIIKGSLDIAPGDMLDKVARRLSLIGHPCSTMSSGSKALEHAKOGNRTHPDIKPPLH 245
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Db      | 211 YOLLGESVDDDAAGEAFDKMTNLIGL-GYP-----GGPETARLERKETPORFYFPRPMTD 263
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Oy      | 246 AKONDSPFGTLQHOYTDCIIMKKKEBEGIERKQILLSAADIAATVOHTMACHLVTRTHRAI 305
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Db      | 264 RRGDLDFPSGLKTFETLN-TWORCVENADDSEQ---TCODIALAFQTAVEETLLIKCRAL 319
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Oy      | 306 LPCKQRDLLPONNAVLYVASGVASNPTIRALTEILTNAOTCTLCPPLLCTDNGINTIAM 365
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Db      | 320 ---KQTL--KN---LVYAGVSANOALRSGLKEMLEGMGKVGFAYARPRTCDNGAMIAY 371
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Oy      | 366 NGIERLARAAGLIHLDIGIRYEPRCPL 392
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Db      | 372 AGCCRILAG--OHDPAISIVOPRWPM 395
        |||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:

RESULT 10
US-08-087-797-3
; Sequence 3, Application US/08087797
; Patent No. 5543312
GENERAL INFORMATION:
APPLICANT: Mellors, Alan
APPLICANT: Lo, Reggie Y.C.
APPLICANT: Abdullah, Khalid M.
TITLE OF INVENTION: Pasteurella Haemolytica
TITLE OF INVENTION: Glycoprotease
```


TELEFAX: 704 334 2014
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 325 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-087-797-2

Query Match 22.4%; Score 475.5; DB 1; Length 325;
 Best Local Similarity 35.8%; Pred. No. 1.5e-45;
 Matches 120; Conservative 47; Mismatches 149; Indels 19; Gaps 6;

QY 39 VLGIETSCDDTAADVDEGTAVLGEAHSQTEVHLKTGGIYVPPAAQOLHRENIQRIVOEA 98
 DB 3 ILGIETSCDEGTAVIYDEDKGVANQLYSQIDMADYGGVPELASHRIKRLTLPLOEA 62
 QY 99 LSASGVSPDSLSAATITIKPGIALSLGVLSPSLQVLQKKPFIPIHMEAHALTIRL- 157
 DB 63 LKEANLPDSIDG:AYTAGPGVGNALVGSITASLAVAMNVPALGVHMECHLAPMLE 122
 QY 158 TNKPEPFLVLLISGCHLALVGVSPFLILGSKSLDIAPGMDLKVARRLSLIHPECS 217
 DB 123 ENAEPEFPVALLISGHTQLVKVGVGQYELIGESIDDAAGEAFDKTGLGL-DYP-- 178
 QY 218 TMSGKALIEHLAKGNRFHDIKPLHAKNCDSPFGLOHTVDKIMKKEEGIEKQ 277
 DB 179 --AGVAMSKAESGTPEPRMTDRPGLDSEFSGLKTPAANTITANINENGBLBO 235
 QY 278 ILSASADIATVQHTMACHLVTRTRAILFCQKQDLFPQNNAVLVSAGVSNFYIRAL 337
 DB 236 ---TKCIIAHAFQAV-----VDIILIKCK-RALEQGYKRLVMAGVGSANKQLRADL 284
 QY 338 EILTNAQCTLLCPPRLCTDNGIMIANGIERLR 372
 DB 285 AEMMKIKGEVFPPECTDNGAMIAVGTFLRK 319

RESULT 13
 US-09-489-9221
 ; Sequence 9221, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709, 2004001
 ; CURRENT APPLICATION NUMBER: US/09/489, 039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 9221
 ; LENGTH: 343
 ; TYPE: PR
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-9221

Query Match 22.3%; Score 474.5; DB 4; Length 343;
 Best Local Similarity 34.5%; Pred. No. 2.1e-45;
 Matches 123; Conservative 51; Mismatches 152; Indels 31; Gaps 8;

QY 39 VLGIETSCDDTAADVDEGTAVLGEAHSQTEVHLKTGGIYVPPAAQOLHRENIQRIVOEA 98
 DB 9 VLGIETSCDEGTAVIYDEDKGVANQLYSQIDMADYGGVPELASHRIKRLTLPLOEA 68
 QY 99 LSASGVSPDSLSAATITIKPGIALSLGVLSPSLQVLQKKPFIPIHMEAHALTIRL- 157
 DB 63 LKEANLPDSIDG:AYTAGPGVGNALVGSITASLAVAMNVPALGVHMECHLAPMLE 128
 QY 158 TNKPEPFLVLLISGCHLALVGVSPFLILGSKSLDIAPGMDLKVARRLSLIHPECS 217
 DB 129 DNPAFPFVALLISGHTQLVSVTIGQYELIGESIDDAAGEAFDKTGLGL-DYP--- 184

QY 218 TMSGKALIEHLAKGNRFHDIKPLHAKNCDSPFGLOHTVDKIMKKEEGIEKQ 277
 DB 185 ---GGPMLSKASQTEGRFVPRMTDRPGLDSEFSGLKTPAANTITANINENGBLBO 235
 QY 278 ILSASADIATVQHTMACHLVTRTRAILFCQKQDLFPQNNAVLVSAGVSNFYIRAL 336
 DB 236 --QTRADIARAFEDAVVDITMICKRA-----LEQYGFRLVMAGVGSANKQLRADL 284
 QY 337 EILTNAQCTLLCPPRLCTDNGIMIANGIERLRIRAGIILHDE-GIRYKCPDL 392
 DB 285 LAEMMQKRGVFPVAPPECTDNGAMIAVGNRLQTA---KALGVTRPRWPL 337

RESULT 14
 US-09-107-532A-6609
 ; Sequence 6609, Application US/09107532A
 ; Patent No. 6563275
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSER: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532A
 ; FILING DATE: 30-Jun-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-5007
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 6609:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 363 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (B) LOCATION 1...363
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6609:
 ; US-09-107-532A-6609

Query Match 22.2%; Score 472; DB 4; Length 363;
 Best Local Similarity 35.7%; Pred. No. 4.5e-45;
 Matches 119; Conservative 59; Mismatches 123; Indels 32; Gaps 9;

QY 37 KVLGIETSCDDTAADVDEGTAVLGEAHSQTEVHLKTGGIYVPPAAQOLHRENIQRIVO 96
 DB 29 ELIATIESCDETSVAVVANGTEILSNIVASQINSHKRGGVPEVASRRHVQITLCLE 88

QY 97 EALSASGVSPDLSNAITTIKPGALSLGVLSFSLQVLGQKKPFIHNEAHALTIR 156
 DB 89 DALVEAGSASBDLSAVAVTYPGLVSGILIGISAKAFAMAHQPLIPVNHMAGHIAAR 148
 QY 157 LTNKVEPPLVLLISGHCCLALVGVSDPFLIGKSLDIAPGMDLKVARRSLIKHPEC 216
 DB 149 LVKFPQPLMLLVSGGHTLVVMQEDSYELIGETRDMAGEADVKGVRVGL----- 202
 QY 217 STMSGKAIEHLAKOG-NRPFHDIKPPLHAKNCDPFTGLQHTVDKTIIMK-KEKEEGIE 274
 DB 203 -SYSGEIDDLAOGKDNHYF--PRAMIHEDNYPFSFGLSAATINLVHNAQORGEDID 259
 QY 275 KGQILSSAADIATVQHTMACHLVKTRRAILFCQORDLPQNNAV--LVASGVASNFY 332
 DB 260 KN-----DLASFQASVIDVLTITLRA---C-----QNYVQQLVAVGVAANQG 302
 QY 333 IRRALIELTNA--TQCTILCPPLCTDNGIMI 363
 DB 303 LREGLOALSAKLPVELVLPPLRLCGDNMAAMI 335

RESULT 15

US-08-987-121A-4
 ; Sequence 4, Application US/08987121A
 ; GENERAL INFORMATION:
 ; APPLICANT: Hoskins, Jo Ann
 ; APPLICANT: Tang, Joseph Chiu-Chung
 ; APPLICANT: Treadway, Patci Jean
 ; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
 ; TITLE OF INVENTION: Gcp
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Eli Lilly and Company
 ; STREET: Lilly Corporate Center
 ; CITY: Indianapolis
 ; STATE: Indiana
 ; COUNTRY: U.S.
 ; ZIP: 46285
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/987,121A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Webster, Thomas D.
 ; REGISTRATION NUMBER: 39,872
 ; REFERENCE/DOCKET NUMBER: X-
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 317-276-3334
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 336 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-987-121A-4

Query Match 21.7%; Score 461; DB 3; Length 336;
 Best Local Similarity 34.5%; Pred. No. 7.3e-44;
 Matches 116; Conservative 58; Mismatches 132; Indels 30; Gaps 8;

QY 37 KIVLGIESDDPTAAVVDERTGNVLGRHISOTREHLTKGIVPPAAQLRENIQRIYQ 96
 DB 4 RYLAFTSCDEISVAALKNDDELNSVIAQISHRKFGGVPEVVASRHHVEVITACIE 63
 QY 97 EALSAGCVSGDLSAIAITTIKPGALSLGVLSFSLQVLGQKKPFIHNEAHALTIR 156
 DB 64 EALABAGITBEDVAVAVTYPGLVSGILIGISAKAFAMAHQPLIPVNHMAGHIAAQ 123

QY 157 LTNKVEPPLVLLISGHCCLALVGVSDPFLIGKSLDIAPGMDLKVARRSLIKHPEC 216
 DB 124 SVEPLEPPLIALLVSGGHTLVVVSAGDYKIYGETRDAVGEAYDKVGVMGJ----- 177
 QY 217 STMSGKAIEHLAKOGNRPFHDIKPPLHAKNCDPFTGLQHTVDKTIIMKKEEGIEKG 276
 DB 178 -TYPAGREIDEHLAQCHDI-YDPPRAMIKEDNLEFSPGKSAFINLHNAE-----QKG 230
 QY 277 QILSSAADIATVQHTMACHLVKTRRAILFCQORDLPQNNAVLVASGVASNFYIRRA 336
 DB 231 ESLSL-EDLCASFQAAVMDILMAKTKKAL-----EKYPVK--TLVAVGVAANKGLRER 281
 QY 337 LRLTNATQCTILCPPLCTDNGIMI-----WN 366
 DB 282 L--ATREITDVNVIIPPLRLCGDNAGMIAYASVSEWN 315

Search completed: February 16, 2005, 13:10:24
 Job time : 27.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:08:47 ; Search time 239.85 Seconds
(without alignments)
563.995 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 2125
Sequence: 1 MLILTKTAGVFPKSRKRVY.....DISKVGSAIKVQLKMEI 414

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	US-10-067-443-2	Sequence 2, Appl1
2	2125	100.0	414	US-10-649-273-2	Sequence 2, Appl1
3	2125	100.0	414	US-10-651-722-2	Sequence 2, Appl1
4	2090.5	98.4	414	US-10-067-443-19	Sequence 19, Appl1
5	2090.5	98.4	439	US-10-649-273-19	Sequence 19, Appl1
6	2090.5	98.4	439	US-10-651-722-19	Sequence 19, Appl1
7	2088	98.3	414	US-10-012-140-5	Sequence 5, Appl1
8	1845	86.8	364	US-10-094-749-2039	Sequence 239, Ap
9	1385	65.2	267	US-10-067-443-22	Sequence 22, Appl1
10	1385	65.2	267	US-10-651-722-22	Sequence 22, Appl1
11	681.5	32.1	445	US-10-424-599-209259	Sequence 209259,
12	681.5	32.1	445	US-10-424-599-209259	Sequence 209259,
13	634	29.8	463	US-10-067-443-3	Sequence 3, Appl1

14	634	29.8	463	15	US-10-649-273-3	Sequence 3, Appl1
15	634	29.8	463	16	US-10-651-722-3	Sequence 3, Appl1
16	578	27.2	444	15	US-10-437-963-113732	Sequence 113732,
17	556.5	26.2	333	14	US-10-012-140-25	Sequence 25, Appl1
18	549	25.8	179	14	US-10-067-443-25	Sequence 25, Appl1
19	549	25.8	179	15	US-10-649-273-25	Sequence 25, Appl1
20	549	25.8	179	15	US-10-651-722-25	Sequence 25, Appl1
21	524	24.7	382	15	US-10-282-122A-50858	Sequence 50858, A
22	524	24.7	421	14	US-10-067-443-4	Sequence 4, Appl1
23	524	24.7	421	14	US-10-067-443-28	Sequence 28, Appl1
24	524	24.7	421	15	US-10-649-273-4	Sequence 28, Appl1
25	524	24.7	421	15	US-10-649-273-28	Sequence 28, Appl1
26	524	24.7	421	15	US-10-651-722-4	Sequence 4, Appl1
27	524	24.7	421	15	US-10-651-722-28	Sequence 28, Appl1
28	511.5	24.1	347	14	US-10-012-140-24	Sequence 24, Appl1
29	502	23.6	348	15	US-10-282-122A-63156	Sequence 63156, A
30	494.5	23.3	343	15	US-10-282-122A-67227	Sequence 67227, A
31	492.5	23.2	341	9	US-09-815-242-11798	Sequence 11798, A
32	492.5	23.2	341	15	US-10-282-122A-66200	Sequence 66200, A
33	489.5	23.0	335	15	US-10-282-122A-55404	Sequence 55404, A
34	484.5	22.8	337	15	US-09-815-242-10304	Sequence 10304, A
35	484.5	22.8	337	15	US-10-282-122A-56695	Sequence 56695, A
36	484.5	22.8	337	15	US-10-282-122A-75485	Sequence 75485, A
37	483.5	22.8	337	15	US-10-282-122A-78161	Sequence 78161, A
38	482.5	22.7	340	15	US-10-282-122A-68438	Sequence 68438, A
39	481.5	22.7	337	9	US-09-815-242-11780	Sequence 11780, A
40	475.5	22.4	342	9	US-09-815-242-11043	Sequence 11043, A
41	475.5	22.4	342	15	US-10-282-122A-58204	Sequence 58204, A
42	469	22.1	340	15	US-10-282-122A-52054	Sequence 52054, A
43	468	22.0	338	15	US-10-282-122A-57817	Sequence 57817, A
44	465.5	21.9	354	15	US-10-282-122A-65768	Sequence 65768, A
45	465	21.9	341	15	US-10-282-122A-67993	Sequence 67993, A

ALIGNMENTS

RESULT 1
US-10-067-443-2
; Sequence 2, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT FILING DATE: 2002-02-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-067-443-2

Query Match 100.0%; Score 2125; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLILTKTAGVFPKSRKRVYFLRSFNHPTGLFLHKIVLGITSCDDPTAAVVDENGV 60
DB 1 MLILTKTAGVFPKSRKRVYFLRSFNHPTGLFLHKIVLGITSCDDPTAAVVDENGV 60
QY 61 LGEATHSQTEVHLKKGIVPPAAQQLHRENTORIQEALSSAGVSPSDLSAATITIKPEL 120
DB 61 LGEATHSQTEVHLKKGIVPPAAQQLHRENTORIQEALSSAGVSPSDLSAATITIKPEL 120
QY 121 ALSICGVGSFSLQVLGQLKKPEPIPIHMEAHALITRLTNKVEPPLVILLISGHCILALV 180

Db 121 ALSTGVGSLFSLQVLGQKPKPIPIHMEBAHALTRITNKVEPPLVLLISGHCILALV 180
|||
Qy 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGNRRFPDIX 240
|||
Db 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGNRRFPDIX 240
|||
Qy 241 PRLHAKKCDPSFTGLQHVTDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Db 241 PRLHAKKCDPSFTGLQHVTDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Qy 301 THRAILFCQKQDRLPQNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360
|||
Db 301 THRAILFCQKQDRLPQNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360
|||
Qy 361 IMIANGIERLRAGGILHDIEGIRYBPKPLGVDSISKEVGASIKVPOLKMEI 414
|||
Db 361 IMIANGIERLRAGGILHDIEGIRYBPKPLGVDSISKEVGASIKVPOLKMEI 414
|||
RESULT 2
US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2
Query Match 100.0%; Score 2125; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLILTKTAGVFFPKPSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDBETGV 60
|||
Db 1 MLILTKTAGVFFPKPSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDBETGV 60
|||
Qy 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAATTTIKPGL 120
|||
Db 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAATTTIKPGL 120
|||
Qy 121 ALSTGVGSLFSLQVLGQKPKPIPIHMEBAHALTRITNKVEPPLVLLISGHCILALV 180
|||
Db 121 ALSTGVGSLFSLQVLGQKPKPIPIHMEBAHALTRITNKVEPPLVLLISGHCILALV 180
|||
Qy 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGNRRFPDIX 240
|||
Db 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGNRRFPDIX 240
|||
Qy 241 PRLHAKKCDPSFTGLQHVTDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Db 241 PRLHAKKCDPSFTGLQHVTDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Qy 301 THRAILFCQKQDRLPQNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360
|||
Db 301 THRAILFCQKQDRLPQNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360
|||
Qy 361 IMIANGIERLRAGGILHDIEGIRYBPKPLGVDSISKEVGASIKVPOLKMEI 414
|||

Db 361 IMIANGIERLRAGGILHDIEGIRYBPKPLGVDSISKEVGASIKVPOLKMEI 414
|||
RESULT 3
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2
Query Match 100.0%; Score 2125; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLILTKTAGVFFPKPSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDBETGV 60
|||
Db 1 MLILTKTAGVFFPKPSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDBETGV 60
|||
Qy 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAATTTIKPGL 120
|||
Db 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAATTTIKPGL 120
|||
Qy 121 ALSTGVGSLFSLQVLGQKPKPIPIHMEBAHALTRITNKVEPPLVLLISGHCILALV 180
|||
Db 121 ALSTGVGSLFSLQVLGQKPKPIPIHMEBAHALTRITNKVEPPLVLLISGHCILALV 180
|||
Qy 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGNRRFPDIX 240
|||
Db 181 QGVSDPFLLGKSLDAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKOGNRRFPDIX 240
|||
Qy 241 PRLHAKKCDPSFTGLQHVTDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Db 241 PRLHAKKCDPSFTGLQHVTDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Qy 301 THRAILFCQKQDRLPQNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360
|||
Db 301 THRAILFCQKQDRLPQNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360
|||
Qy 361 IMIANGIERLRAGGILHDIEGIRYBPKPLGVDSISKEVGASIKVPOLKMEI 414
|||
Db 361 IMIANGIERLRAGGILHDIEGIRYBPKPLGVDSISKEVGASIKVPOLKMEI 414
|||
RESULT 4
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518

;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 60/282,814
;; PRIOR FILING DATE: 2001-04-10
;; NUMBER OF SEQ ID NOS: 71
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 19
;; LENGTH: 439
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-067-443-19

Query Match 98.4%; Score 2090.5; DB 14; Length 439;
Best Local Similarity 93.8%; Pred. No. 4e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFPFKSKRYVEFLRSFNHFGTLFLHKIVLGIETSCDDPTAAAVVDEGTNV 60
DB 1 MLITKTAGVFPFKSKRYVEFLRSFNHFGTLFLHKIVLGIETSCDDPTAAAVVDEGTNV 60
QY 61 LGEAIIHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPDLSAIAITIKPGL 120
DB 61 LGEAIIHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPDLSAIAITIKPGL 120
QY 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALTIRLTNKEVFPFLVLLISGGHCLLALV 180
DB 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALTIRLTNKEVFPFLVLLISGGHCLLALV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECTWSGGKAIIEHLAKQGRHFPIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECTWSGGKAIIEHLAKQGRHFPIK 240
QY 241 PPLHAKNCDFSTGLQHVTDKTIIMKKEKEGFI-----EK 275
DB 241 PPLHAKNCDFSTGLQHVTDKTIIMKKEKEGFI-----EK 275
QY 276 GQILSSADIAATVOHTMACHLVKRTIRAILFCQKRDLLPONNAVLVASGVASNFYIR 335
DB 301 GQILSSADIAATVOHTMACHLVKRTIRAILFCQKRDLLPONNAVLVASGVASNFYIR 360
QY 336 ALBITTNAOTCTLLCPPRLCTDNGIMIAMNGIERLRAGLGIILHDEGIRYEPKCPGLVD 395
DB 361 ALBITTNAOTCTLLCPPRLCTDNGIMIAMNGIERLRAGLGIILHDEGIRYEPKCPGLVD 420
QY 396 ISKEVGEASIKVPOLKMEI 414
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 5
US-10-649-273-19
;; Sequence 19, Application US/10649273
;; Publication No. US20040043407A1
;; GENERAL INFORMATION:
;; APPLICANT: Bristol-Myers Squibb Company
;; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
;; FILE REFERENCE: D0073 CNT
;; CURRENT APPLICATION NUMBER: US/10/649,273
;; PRIOR FILING DATE: 2003-08-27
;; PRIOR APPLICATION NUMBER: US 60/266,518
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 10/067,443
;; PRIOR FILING DATE: 2002-02-05
;; PRIOR APPLICATION NUMBER: US 60/282,814
;; PRIOR FILING DATE: 2001-04-10
;; NUMBER OF SEQ ID NOS: 71
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 19
;; LENGTH: 439
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-649-273-19

Query Match 98.4%; Score 2090.5; DB 15; Length 439;

Best Local Similarity 93.8%; Pred. No. 4e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFPFKSKRYVEFLRSFNHFGTLFLHKIVLGIETSCDDPTAAAVVDEGTNV 60
DB 1 MLITKTAGVFPFKSKRYVEFLRSFNHFGTLFLHKIVLGIETSCDDPTAAAVVDEGTNV 60
QY 61 LGEAIIHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPDLSAIAITIKPGL 120
DB 61 LGEAIIHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPDLSAIAITIKPGL 120
QY 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALTIRLTNKEVFPFLVLLISGGHCLLALV 180
DB 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALTIRLTNKEVFPFLVLLISGGHCLLALV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECTWSGGKAIIEHLAKQGRHFPIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECTWSGGKAIIEHLAKQGRHFPIK 240
QY 241 PPLHAKNCDFSTGLQHVTDKTIIMKKEKEGFI-----EK 275
DB 241 PPLHAKNCDFSTGLQHVTDKTIIMKKEKEGFI-----EK 275
QY 276 GQILSSADIAATVOHTMACHLVKRTIRAILFCQKRDLLPONNAVLVASGVASNFYIR 335
DB 301 GQILSSADIAATVOHTMACHLVKRTIRAILFCQKRDLLPONNAVLVASGVASNFYIR 360
QY 336 ALBITTNAOTCTLLCPPRLCTDNGIMIAMNGIERLRAGLGIILHDEGIRYEPKCPGLVD 395
DB 361 ALBITTNAOTCTLLCPPRLCTDNGIMIAMNGIERLRAGLGIILHDEGIRYEPKCPGLVD 420
QY 396 ISKEVGEASIKVPOLKMEI 414
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 6
US-10-651-722-19
;; Sequence 19, Application US/10651722
;; Publication No. US20040048302A1
;; GENERAL INFORMATION:
;; APPLICANT: Bristol-Myers Squibb Company
;; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
;; FILE REFERENCE: D0073 DIV
;; CURRENT APPLICATION NUMBER: US/10/651,722
;; PRIOR FILING DATE: 2003-08-29
;; PRIOR APPLICATION NUMBER: US 60/266,518
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 10/067,443
;; PRIOR FILING DATE: 2002-02-05
;; PRIOR APPLICATION NUMBER: US 60/282,814
;; PRIOR FILING DATE: 2001-04-10
;; NUMBER OF SEQ ID NOS: 71
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 19
;; LENGTH: 439
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-651-722-19

Query Match 98.4%; Score 2090.5; DB 15; Length 439;
Best Local Similarity 93.8%; Pred. No. 4e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFPFKSKRYVEFLRSFNHFGTLFLHKIVLGIETSCDDPTAAAVVDEGTNV 60
DB 1 MLITKTAGVFPFKSKRYVEFLRSFNHFGTLFLHKIVLGIETSCDDPTAAAVVDEGTNV 60
QY 61 LGEAIIHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPDLSAIAITIKPGL 120
DB 61 LGEAIIHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALASAGVSPDLSAIAITIKPGL 120
QY 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALTIRLTNKEVFPFLVLLISGGHCLLALV 180

Db	121	ALSLGVGSLFSLQVLGQLKKPPIPIHMHBAHLTRITKNVPEPFLVLLISGHCLIALV	180
QY	181	QGVSDPFLILGSLDIAPEGMDLKVARRSLIKHPECSNMSGKAIIEHLAKQGNRPFFDIK	240
Db	181	QGVSDPFLILGSLDIAPEGMDLKVARRSLIKHPECSNMSGKAIIEHLAKQGNRPFFDIK	240
QY	241	PLHMHAKNCDSFPGLOHYTDKIKKKKEEGI-----EK	275
Db	241	PLHMHAKNCDSFPGLOHYTDKIKKKKEEGIPLISKVEQINIPEGLCKIAHPCRYEK	300
QY	276	GQLISSAADIAATVQHTACHLVKEKTHRAILFCQKRDLLPONNAVVLVASGVASNFYIRR	335
Db	301	GQLISSAADIAATVQHTACHLVKEKTHRAILFCQKRDLLPONNAVVLVASGVASNFYIRR	360
QY	336	ALBLITNAQTCTLLCPPRLCTDNGIMTAMNGIERIRIAGIILHDIIEGIRYEPKCPVLGVD	395
Db	361	ALBLITNAQTCTLLCPPRLCTDNGIMTAMNGIERIRIAGIILHDIIEGIRYEPKCPVLGVD	420
QY	396	ISKVEGASIRVPOLKMEI	414
Db	421	ISKVEGASIRVPOLKMEI	439

```

RESULT 7
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US2003009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 38152004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

Query Match          98.3%; Score 2088; DB 14; Length 414;
Best local Similarity 98.3%; Pred. No. 6, 4e-196;
Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 MLILTKTAGVFPKPSKRYVEFLRSFNFPQTLFLHKIVLGIETSCDDTAAYVDEGTNV 60
Db      1 MLILTKTAGVFPKPSKRYVEFLRSFNFPQTLFLHKIVLGIETSCDDTAAYVDEGTNV 60

QY      61 LGEAHSSTEVYHLTKGTGIYPPAAQQLHRENIORIYOEALSASGVSPSDLSAINTTKPGL 120
Db      61 LGEAHSSTEVYHLTKGTGIYPPAAQQLHRENIORIYOEALSASGVSPSDLSAINTTKPGL 120

QY      121 ALISGVGVSTLQVNGOLKKPFIPIHMEAHALTLRLTNKVEPFLVLLISGHCCLALY 180
Db      121 ALISGVGVSTLQVNGOLKKPFIPIHMEAHALTLRLTNKVEPFLVLLISGHCCLALY 180

QY      181 QGVSDPFLILGKSLDAPGMDLKVARRSLIIGHPECTSSGGKAIETHLAKQGRFFPDIK 240
Db      181 QGVSDPFLILGKSLDAPGMDLKVARRSLIIGHPECTSSGGKAIETHLAKQGRFFPDIK 240

QY      241 PPLHAAKNCDFSTGLQVHTDKIINKKEKEGCIKQQLISSAADIAATVQHTMACHLVKR 300
Db      241 PPLHAAKNCDFSTGLQVHTDKIINKKEKEGCIKQQLISSAADIAATVQHTMACHLVKR 300

```

[illegible]

```

RESULT 8
US-10-094-749-2039
/ Sequence 2039, Application US/10094749
/ Publication No. US20030219741A1
/ GENERAL INFORMATION:
/ APPLICANT: ISOGAI, TAKAO
/ APPLICANT: SUGIYAMA, TOMOYASU
/ APPLICANT: OTSUKI, TETSUJI
/ APPLICANT: WAKAMATSU, AI
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUUKO
/ APPLICANT: HIO, YURI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: IRIE, RYOTARO
/ APPLICANT: TAMECHIKA, ICHIRO
/ APPLICANT: SEKI, NAOHICO
/ APPLICANT: YOSHIKAWA, TSUTOMU
/ APPLICANT: OTSUKA, MOTOFUKI
/ APPLICANT: NAGAHARI, KENJI
/ APPLICANT: MASUHO, YASUHIKO
/ TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
/ FILE REFERENCE: 084335/0160
/ CURRENT APPLICATION NUMBER: US/10/094,749
/ CURRENT FILING DATE: 2002-03-12
/ PRIOR APPLICATION NUMBER: 60/350,435
/ PRIOR FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: JP 2001-328381
/ PRIOR FILING DATE: 2001-09-14
/ NUMBER OF SEQ ID NOS: 3381
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2039
/ LENGTH: 364
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-094-749-2039

Query Match 86.8%; Score 1845; DB 15; Length 364;
Best Local Similarity 99.2%; Pred. No. 3.8e-172;
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLILITAGVFFPKSKRYVEFLRSFNFHPTGLFLAKIVLGIETSCDDTAAAVDEITGV 60
Db 1 MLILITAGVFFPKSKRYVEFLRSFNFHPTGLFLAKIVLGIETSCDDTAAAVDEITGV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAOOLHRENIQRIVOELASGVSPSLAATTIKXGL 120
Db 61 LGEAHSQTEVHLKTGGIVPPAAOOLHRENIQRIVOELASGVSPSLAATTIKXGL 120
QY 121 ALSTGVGTSFSLQVGOLKKEPFIPIHMEAHALTIRLTNKEFPEPLVLLISGHCLALV 180
Db 121 ALSTGVGTSFSLQVGOLKKEPFIPIHMEAHALTIRLTNKEFPEPLVLLISGHCLALV 180
QY 181 QGVSDPFLILGSLDIAPGDMLDKVARRLSLIKHPECSTMSGGKAIETHLAKQGNRFHFDIX 240
Db 181 QGVSDPFLILGSLDIAPGDMLDKVARRLPLIKHPECSTMSGGKAIETHLAKQGNRFHFDIX 240
QY 241 PRLHAAKCDPSFPGLOHTVTDKTIIMCKEKEGIEKQGLSSADIAATVQHTMACHLVYK 300
Db 241 PRLHAAKCDPSFPGLOHTVTDKTIIMCKEKEGIEKQGLSSADIAATVQHTMACHLVYK 300

```

Db 241 PRLHAKNDFPFTGLQHTVDKIIMKEKEGEGKQILSSADIAATVQHTMACHLVKR 300
Qy 301 THRAILFCQKORDLLPQNNNAVIVASGVSANFYIRALEILTNATQCTLLCPPRLCTDNG 360
Db 301 THRAILFCQKORDLLPQNNNAVIVASGVSANFYIRALEILTNATQCTLLCPPRLCTDNG 360
Qy 361 IMA 364
Db 361 IMA 364

RESULT 9

US-10-674-443-22
; Sequence 22, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-674-443-22

Query Match 65.2%; Score 1385; DB 14; Length 267;
Best Local Similarity 100.0%; Pred. No. 3.5e-127;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 MEAALTLRLTNKVEPPPLVLLISGHCILALVQVSDFLGKSLDIAPGMDLVARR 207
Db 1 MEAALTLRLTNKVEPPPLVLLISGHCILALVQVSDFLGKSLDIAPGMDLVARR 60
Qy 208 LSLIKHPECSMTSGGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLQHTVDKIIMK 267
Db 61 LSLIKHPECSMTSGGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLQHTVDKIIMK 120
Qy 268 EKEGIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 327
Db 121 EKEGIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 180
Qy 328 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGLHDIGIRYE 387
Db 181 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGLHDIGIRYE 240
Qy 388 PKCPLGVDISKVEGASIKVPOLKMEI 414
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 10

US-10-649-273-22
; Sequence 22, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-22

Query Match 65.2%; Score 1385; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 3.5e-127;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 MEAALTLRLTNKVEPPPLVLLISGHCILALVQVSDFLGKSLDIAPGMDLVARR 207
Db 1 MEAALTLRLTNKVEPPPLVLLISGHCILALVQVSDFLGKSLDIAPGMDLVARR 60
Qy 208 LSLIKHPECSMTSGGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLQHTVDKIIMK 267
Db 61 LSLIKHPECSMTSGGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLQHTVDKIIMK 120
Qy 268 EKEGIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 327
Db 121 EKEGIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 180
Qy 328 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGLHDIGIRYE 387
Db 181 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGLHDIGIRYE 240
Qy 388 PKCPLGVDISKVEGASIKVPOLKMEI 414
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 11

US-10-651-722-22
; Sequence 22, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-22

Query Match 65.2%; Score 1385; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 3.5e-127;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 MEAALTLRLTNKVEPPPLVLLISGHCILALVQVSDFLGKSLDIAPGMDLVARR 207
Db 1 MEAALTLRLTNKVEPPPLVLLISGHCILALVQVSDFLGKSLDIAPGMDLVARR 60
Qy 208 LSLIKHPECSMTSGGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLQHTVDKIIMK 267
Db 61 LSLIKHPECSMTSGGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLQHTVDKIIMK 120
Qy 268 EKEGIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 327
Db 268 EKEGIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 327

Db 121 EKEEGIEKQILSSAADIATVQHTMACHLVKTRRAILFCQKORDLPQNNAVLVASGV 180
Qy 328 ASNFYIRALIELTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 387
Db 181 ASNFYIRALIELTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 240
Qy 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414
Db 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267

RESULT 12

US-10-424-599-209259
; Sequence 209259, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209259
; LENGTH: 445
; TYPE: PRP
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3098C.1.pep
US-10-424-599-209259

Query Match 32.1%; Score 681.5; DB 15; Length 445;
Best Local Similarity 43.1%; Pred. No. 8,8e-58;
Matches 155; Conservative 54; Mismatches 126; Indels 25; Gaps 6;

Qy 38 IVLGIEISCDTAAAVDETVNGVLEAHSQTEVHLKTGIVPPAAQOLHRENIQRIYOE 97
Db 55 VVLGIEISCDTAAAVVSDDEILSQVSSQADILAKTGVA PKAEASHSVIDQVVOE 114
Qy 98 ALSASGVSPDLSAATATTKPGLALSLGVSFSLQVQLKKPFIPIHMEHALTIRL 157
Db 115 ALDRAVLTEDQLTAVAVTIGPGLSLCLRVGVOKAKIKGPTLPITIGHMEHALVRL 174
Qy 158 TNK-VEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVARRLSLIGHEPC 216
Db 175 IEKLOPPFMAILLISGHNLLVLARDLGQYIQLGTTIDDAIGEAYDKTAKVLGLRL-- 231
Qy 217 STMSGKALIEHLAKQGRFHFIDIKPPLHAKKCDSPFTGLQ----HYTDKIMKEKEE 271
Db 232 --RSGGAIEKTLAEGNAESVKSIFPMKQHKDCNFSYAGLKTQVRLAESKKIDAKIPIIS 289
Qy 272 GIEKGQILSSAADIATVQHTMACHLVKTRRAILFCQKORDLPQNNAVLVASGVASNF 331
Db 290 SASNGDEL-SPADTAAASFQRAVLAHLERCEBAIDMALKEPSIH---LVVSGVASNQ 345
Qy 332 YIRRALIELTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYEPKCP 391
Db 346 YVRRALIMVVKNGQLQVCPPLRLCTDNGVMIAMGIEHFRNG-----RYDPPPP 395

RESULT 13

US-10-067-443-3
; Sequence 3, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443

; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRP
; ORGANISM: Arabidopsis thaliana
US-10-067-443-3

Query Match 29.8%; Score 634; DB 14; Length 463;
Best Local Similarity 37.9%; Pred. No. 4,3e-53;
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

Qy 38 IVLGIEISCDTAAAVDETVNGVLEAHSQTEVHLKTGIVPPAAQOLHRENIQRIYOE 97
Db 85 VVLGIEISCDTAAAVVSPFNHSSC---RAELLVQGVAPAKQAEASHSVIDKVVOD 141
Qy 98 ALSASGVSPDLSAATATTKPGLALSLGVSFSLQVQLKKPFIPIHMEHALTIRL 157
Db 142 ALDKANLTREKDSAAVAVTIGPGLSLCLRVGAKARVAGNFSLPYGVHMEHALVRL 201
Qy 158 T-NKVEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVARRLSLIGHEPC 216
Db 202 VQGLSFPFMAILLISGHNLLVLAKHGLQYIQLGTTIDDAIGEAFKTKAKVLGLDNH--- 258
Qy 217 STMSGKALIEHLAKQGRFHFIDIKPPLHAKKCDSPFTGLQHYTDKIMKEKEEGIEKG 276
Db 259 --RSGGPAVEBELADGASVAFNPMKVHKDCNFSYAGLKTQVRLAIEKX----- 308
Qy 277 QILSSAADIATVQHTMACHLVKTRRAILFCQKORDLPQNNAVLVASGVASNFIRRA 336
Db 309 --IRNRADTAAASFQRAVLAHLERCEBAIDMALE--LEPSIKHNVISGVASNKYVRLR 363
Qy 337 IEILTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE----- 387
Db 364 LNNIVENKMLKAVCPPLSLCTDNGVAVWATGJLHFRVQ-----RYDPPPATPEE 413
Qy 388 -----PKCPLGVDISKEVGEA 403
Db 414 DYVDLRRPWRPLGEBYAKGRSEA 436

RESULT 14

US-10-649-273-3
; Sequence 3, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRP
; ORGANISM: Arabidopsis thaliana
US-10-649-273-3

Query Match 29.8%; Score 634; DB 15; Length 463;
Best Local Similarity 37.9%; Pred. No. 4,3e-53;
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

```

QY 38 IVLGIETSCDDPAAAVNBETGVNLGSAIHSGTEVHLKTGGIIVPPAAQOLRENIORIVQE 97
Dv 85 VVLGIETSCDDPRAAVNSPNNHLSSEC---PAELVQYGVCAKQAEHARSITDKVYD 144
QY 98 ALSASGVSPSDISAIATTTIKPGLASLGVGLSFLQVLGOLKKPPIPIHMEAHATTRL 157
Dv 142 ALDKANTITEKDISAVALTIGPGLSICLRGVKARARVAGNFSPIYGVHMEAHAVARL 201
QY 158 T-NKVBEPFVULLIISGHCGLTALVGVSPFLLIGSLDIAFGMDUKVARLSLIIKPEC 216
Dv 202 VQEGISFPFMALLISGGHNLVLAHKRLGGYDGLGTVDDAIGAEFRTAKMLGIDMH--- 258
QY 217 STMSGKKAIENHLAKQGNRFHDIKPIHLHAKQCDPSFTGJOHYTDTKIIMKCEKEGIEKG 276
Dv 259 --RSGPRAVEBELATSGDAKSQVFNPMKHKQCNFSAIGLKTQVRLAIEAKE----- 308
QY 277 QILSADIAATVOHTMACHLVKTHTRAILFFCKORDLLEPQNNAVLVAASGVASNFYIRRA 336
Dv 309 --IRNRADIAASFQVAVVLHEBEKERALDWALE--LEPSIGHWVISGGVASNKTYRLR 363
QY 337 LEILTNAOCTLLCEPPRLCTNGIMIANGLERLPAIGLILHDIGIRYE----- 387
Dv 364 LNNIIVENKNLKLIVCEPPSLCTNGVWVAWVTGLEHFEFVG-----RYDPPPEATEPE 413
QY 388 -----PKCPLGVDSKEVGEA 403
Dv 414 DYVDIDRFRWPLAGEEYAKGRSEA 436

```

```

RESULT 15
US-10-651-722-3
? Sequence 3, Application US/10651722
? Publication No. US20040048302A1
? GENERAL INFORMATION:
? APPLICANT: Bristol-Myers Squibb Company
? TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
? FILE REFERENCE: D0073 DIV
? CURRENT APPLICATION NUMBER: US/10/651,722
? CURRENT FILING DATE: 2003-08-29
? PRIOR APPLICATION NUMBER: US 60/566,518
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 10/067,443
? PRIOR FILING DATE: 2002-02-05
? PRIOR APPLICATION NUMBER: US 60/282,814
? PRIOR FILING DATE: 2001-04-10
? NUMBER OF SEQ ID NOS: 71
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 3
? LENGTH: 463
? TYPE: PRT
? ORGANISM: Arabidopsis thaliana
? US-10-651-722-3

```

Query Match	29.8%;	Score 634;	DB 15;	Length 463;
Best Local Similarity	37.9%;	Pred. No. 4.3e-53;		
Matches 145; Conservative	61;	Mismatches 129;	Indels 48;	Gaps 7

```

QY      38  IVLGIEISCDTAAAVVDGETGVVGLAEALHSQTEVHLKTGCIYPAQAQQLREHIOGIVOE  97
DQ      85  VVLGIEISCDTAAAVVSPFNHTSSCC---RKLVLQYGVGAPKQAEBAHSRIYDKVQD  141
QY      98  ALSASGVSPDLSAIAVTTIKPGLALSLGVGLSPSLQVGLQKKPPIPIHMHMAHALTIRL  157
DQ     142  ALDKANLTKEKDSAAVAVTITGPELSICLRGVARKARVAGNFSPIYGVHMHMAHALVARL  201
QY     158  T-ANKVEPFPVLLISGGHCLLALVGVSDFFLLAGSLDIAPGMDLKVARRSLIKHPBC  216
DQ     202  VEOEISPPFPMLLISGGHLLVLAHKLCAQYQLGTVVDALISAPAPKTAKTWGLIDNH---  258
QY     217  STMSGKRIEHLAKQGNRFHPDIKPYLHNAKCDSPFGICQHTVQDIKKKEKEBEOIEKG  276
DQ     259  --RSQGRPVVEELAEIGDASKYKENVPMKYHCKDQCNISVALKQVRLAIRAK-----  308

```

```

OY 277 QILSSAAIATAYVOTHTMACHUKVRTHRAIILFCOKRDLLPQNNNAVAVSGVSNPFYTRRA 336
Db 309 --IRRAIDIASFOQVAVALHEEKKERERIALDWALE--LEPSIKHMYISGVASNKYVRLR 363
OY 337 LEIITNATOCTLLCPRPRLCTDNGIMIAMNGIERLRAGLILHDIEGIRYE----- 387
Db 364 LNNIVENGNLKLVCPPSLCTDNGMMVAMVATGLEHFRVG-----RIDPPPPATEPE 413
OY 388 -----KCPGLGVDISKEVEGA 403
Db 414 DYVYDLRPRWPLGEEYAKGRSEA 436

```

Search completed: February 16, 2005, 13:28:29
Job time : 242.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:02:52 ; Search time 226.8 Seconds
(without alignments)
2132.075 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125
Sequence: 1 MLTLTAVGVFPKSRKRVY.....DISKEVGEASIKVPLKMEI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 6959266 seqs, 116806243 residues

Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Pending Patents AA Main:*
- 1: /cgn2_6/prodata/1/paa/PCYUS COMB.pep.*
 - 2: /cgn2_6/prodata/1/paa/US06 COMB.pep.*
 - 3: /cgn2_6/prodata/1/paa/US07 COMB.pep.*
 - 4: /cgn2_6/prodata/1/paa/US08 COMB.pep.*
 - 5: /cgn2_6/prodata/1/paa/US081 COMB.pep.*
 - 6: /cgn2_6/prodata/1/paa/US082 COMB.pep.*
 - 7: /cgn2_6/prodata/1/paa/US083 COMB.pep.*
 - 8: /cgn2_6/prodata/1/paa/US084 COMB.pep.*
 - 9: /cgn2_6/prodata/1/paa/US085 COMB.pep.*
 - 10: /cgn2_6/prodata/1/paa/US086 COMB.pep.*
 - 11: /cgn2_6/prodata/1/paa/US087 COMB.pep.*
 - 12: /cgn2_6/prodata/1/paa/US088 COMB.pep.*
 - 13: /cgn2_6/prodata/1/paa/US089 COMB.pep.*
 - 14: /cgn2_6/prodata/1/paa/US090 COMB.pep.*
 - 15: /cgn2_6/prodata/1/paa/US091 COMB.pep.*
 - 16: /cgn2_6/prodata/1/paa/US092 COMB.pep.*
 - 17: /cgn2_6/prodata/1/paa/US093 COMB.pep.*
 - 18: /cgn2_6/prodata/1/paa/US094 COMB.pep.*
 - 19: /cgn2_6/prodata/1/paa/US095 COMB.pep.*
 - 20: /cgn2_6/prodata/1/paa/US096 COMB.pep.*
 - 21: /cgn2_6/prodata/1/paa/US097A COMB.pep.*
 - 22: /cgn2_6/prodata/1/paa/US097B COMB.pep.*
 - 23: /cgn2_6/prodata/1/paa/US098 COMB.pep.*
 - 24: /cgn2_6/prodata/1/paa/US099A COMB.pep.*
 - 25: /cgn2_6/prodata/1/paa/US099B COMB.pep.*
 - 26: /cgn2_6/prodata/1/paa/US100 COMB.pep.*
 - 27: /cgn2_6/prodata/1/paa/US101 COMB.pep.*
 - 28: /cgn2_6/prodata/1/paa/US102 COMB.pep.*
 - 29: /cgn2_6/prodata/1/paa/US103 COMB.pep.*
 - 30: /cgn2_6/prodata/1/paa/US104 COMB.pep.*
 - 31: /cgn2_6/prodata/1/paa/US105 COMB.pep.*
 - 32: /cgn2_6/prodata/1/paa/US106 COMB.pep.*
 - 33: /cgn2_6/prodata/1/paa/US107 COMB.pep.*
 - 34: /cgn2_6/prodata/1/paa/US108 COMB.pep.*
 - 35: /cgn2_6/prodata/1/paa/US109 COMB.pep.*
 - 36: /cgn2_6/prodata/1/paa/US110 COMB.pep.*
 - 37: /cgn2_6/prodata/1/paa/US60 COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	414	1 PCT-US02-03353-2	Sequence 2, Appl1
2	2125	100.0	414	1 PCT-US02-19360-8	Sequence 8, Appl1
3	2125	100.0	414	30 US-10-480-988-8	Sequence 2, Appl1
4	2125	100.0	414	32 US-10-649-273-2	Sequence 2, Appl1
5	2125	100.0	414	32 US-10-651-722-2	Sequence 2, Appl1
6	2125	100.0	414	32 US-60-266-518-2	Sequence 2, Appl1
7	2125	100.0	414	37 US-60-282-814-2	Sequence 2, Appl1
8	2125	100.0	425	35 US-10-918-754-1372	Sequence 1372, Ap
9	2125	100.0	425	37 US-60-495-114-1372	Sequence 1372, Ap
10	2090.5	98.4	439	1 PCT-US02-03353-19	Sequence 19, Appl
11	2090.5	98.4	439	32 US-10-651-722-19	Sequence 19, Appl
12	2090.5	98.4	439	32 US-10-651-722-19	Sequence 19, Appl
13	2090.5	98.4	439	37 US-60-266-518-19	Sequence 19, Appl
14	2090.5	98.4	439	37 US-60-282-814-19	Sequence 19, Appl
15	2088	98.3	414	26 US-10-012-140-5	Sequence 5, Appl1
16	1845	86.8	364	26 US-10-094-749-2039	Sequence 2039, Ap
17	1395	65.6	298	27 US-10-170-205E-27317	Sequence 27317, A
18	1395	65.6	298	35 US-10-918-754-1371	Sequence 1371, Ap
19	1395	65.6	298	37 US-60-495-114-1371	Sequence 1371, Ap
20	1395	65.6	309	35 US-10-918-754-1374	Sequence 1374, Ap
21	1395	65.6	309	35 US-10-918-754-1375	Sequence 1375, Ap
22	1395	65.6	309	37 US-60-495-114-1374	Sequence 1374, Ap
23	1395	65.6	309	37 US-60-495-114-1375	Sequence 1375, Ap
24	1385	65.2	267	1 PCT-US02-03353-22	Sequence 22, Appl
25	1385	65.2	267	32 US-10-649-273-22	Sequence 22, Appl
26	1385	65.2	267	35 US-10-651-722-22	Sequence 22, Appl
27	1385	65.2	267	37 US-10-918-754-1373	Sequence 1373, Ap
28	1385	65.2	267	37 US-60-266-518-22	Sequence 22, Appl
29	1385	65.2	267	37 US-60-282-814-22	Sequence 22, Appl
30	1385	65.2	267	37 US-60-495-114-1373	Sequence 1373, Ap
31	827	38.9	291	37 US-60-243-468-1011	Sequence 1011, Ap
32	827	38.9	291	37 US-60-243-742-177	Sequence 177, App
33	714.5	33.6	408	37 US-60-173-464-26192	Sequence 26192, A
34	714.5	33.6	409	20 US-09-614-150-34191	Sequence 34191, A
35	714.5	33.6	409	20 US-09-614-150A-34191	Sequence 34191, A
36	714.5	33.6	409	37 US-60-191-637-33775	Sequence 33775, A
37	714.5	33.6	409	37 US-60-191-681-26629	Sequence 26629, A
38	681.5	32.1	445	30 US-10-424-599-209259	Sequence 209259, A
39	659.5	31.0	439	19 US-09-513-996A-21029	Sequence 21029, A
40	659.5	31.0	444	19 US-09-513-996A-21029	Sequence 21029, A
41	656	30.9	461	10 US-10-449-902-48987	Sequence 48987, A
42	634	29.8	463	1 PCT-US02-03353-3	Sequence 3, Appl1
43	634	29.8	463	21 US-09-708-427-16627	Sequence 16627, A
44	634	29.8	463	32 US-10-649-273-3	Sequence 3, Appl1
45	634	29.8	463	32 US-10-651-722-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
PCT-US02-03353-2
Sequence 2, Application PC/TUS0203353
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYPEPTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
FILE REFERENCE: D0073 PCT
CURRENT APPLICATION NUMBER: PCT/US02/03353
PRIORITY FILING DATE: 2002-02-05
PRIORITY FILING DATE: 2001-02-05
PRIORITY FILING DATE: 2001-02-05
PRIORITY FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SEQUENCE: Patent version 3.0
LENGTH: 414
TYPE: PRT
ORGANISM: Homo sapiens

PCT-US02-03353-2

Query Match 100.0%; Score 2125; DB 1; Length 414;
 Best Local Similarity 100.0%; Pred. No. 1.2e-214;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLITKTAGVFPKPSKRKYEFRLSPNFHPTGLFLHKIVLGIFTSCTDDTAAAVVDSTGVN 60
DB 1 MLITKTAGVFPKPSKRKYEFRLSPNFHPTGLFLHKIVLGIFTSCTDDTAAAVVDSTGVN 60
QY 61 LGEAHSQTEVHLKGTGIVPPAAQQLHRENIQRIYQEAALSASGVSPSLSAIAITIKPGL 120
DB 61 LGEAHSQTEVHLKGTGIVPPAAQQLHRENIQRIYQEAALSASGVSPSLSAIAITIKPGL 120
QY 121 ALSGVGSPSLQVGVQLKKPFIPIHMEAHALTIRLTNKVEPPLVLLISGHCILLAV 180
DB 121 ALSGVGSPSLQVGVQLKKPFIPIHMEAHALTIRLTNKVEPPLVLLISGHCILLAV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDFSTGLQHTVDKIIKKKEKEGIEKGQILSSAADIATVQHTMACHLYR 300
DB 241 PPLHAKNCDFSTGLQHTVDKIIKKKEKEGIEKGQILSSAADIATVQHTMACHLYR 300
QY 301 THRALIFCKQRDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
DB 301 THRALIFCKQRDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGIGILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414
DB 361 IMIANGIERLRAGIGILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414

```

RESULT 2

PCT-US02-19360-8

Sequence 8, Application PC/TUS0219360

GENERAL INFORMATION:

```

; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: KABLE, Amy E.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: HARALIA, April J.A.
; APPLICANT: TRAN, Bao
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: ISON, Craig H.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: NGUYEN, Daniel B.
; APPLICANT: LU, Dzung Aina M.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: YU, Henry
; APPLICANT: FORSYTHE, Ian J.
; APPLICANT: BARROSO, Ines
; APPLICANT: RAMKIMAR, Jayalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: LI, Joana X.
; APPLICANT: YANG, Junning
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: DING, Li
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: YAO, Monique G.
; APPLICANT: WALIA, Narinder K.
; APPLICANT: MASON, Patricia M.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: LEE, Sally
; APPLICANT: BECHA, Shanya D.
; APPLICANT: LEE, Soo Yeun
; APPLICANT: TRAN, Uyen K.
; APPLICANT: ELLIOTT, Vicki S.

```

```

; APPLICANT: LUO, Wen
; APPLICANT: SPRAGUE, William
; APPLICANT: TANG, Y. Tom
; APPLICANT: LU, Yan
; APPLICANT: ZEBARADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/19360
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7632424CD1
PCT-US02-19360-8

```

Query Match 100.0%; Score 2125; DB 1; Length 414;
 Best Local Similarity 100.0%; Pred. No. 1.2e-214;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLITKTAGVFPKPSKRKYEFRLSPNFHPTGLFLHKIVLGIFTSCTDDTAAAVVDSTGVN 60
DB 1 MLITKTAGVFPKPSKRKYEFRLSPNFHPTGLFLHKIVLGIFTSCTDDTAAAVVDSTGVN 60
QY 61 LGEAHSQTEVHLKGTGIVPPAAQQLHRENIQRIYQEAALSASGVSPSLSAIAITIKPGL 120
DB 61 LGEAHSQTEVHLKGTGIVPPAAQQLHRENIQRIYQEAALSASGVSPSLSAIAITIKPGL 120
QY 121 ALSGVGSPSLQVGVQLKKPFIPIHMEAHALTIRLTNKVEPPLVLLISGHCILLAV 180
DB 121 ALSGVGSPSLQVGVQLKKPFIPIHMEAHALTIRLTNKVEPPLVLLISGHCILLAV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDFSTGLQHTVDKIIKKKEKEGIEKGQILSSAADIATVQHTMACHLYR 300
DB 241 PPLHAKNCDFSTGLQHTVDKIIKKKEKEGIEKGQILSSAADIATVQHTMACHLYR 300
QY 301 THRALIFCKQRDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
DB 301 THRALIFCKQRDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGIGILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414
DB 361 IMIANGIERLRAGIGILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414

```

RESULT 3

US-10-480-988-8

Sequence 8, Application US/10480988

GENERAL INFORMATION:

```

; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HARALIA, April J.A.;

```



```

APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
APPLICANT: LU, Dzung Alma M.; LEE, Ernestine A.;
APPLICANT: YUE, Henry; FOSSYTHE, Ian U.;
APPLICANT: BARROSO, Ines; RAMKUMAR, Jeyalaxmi;
APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;
APPLICANT: YANG, Junming; THANGAVALU, Kavitha;
APPLICANT: GIETZEN, Kimberly J.; DING, Li;
APPLICANT: BAUGHN, Mariah R.; BOKUMSKY, Mark L.;
APPLICANT: YAO, Montique G.; CHAWLA, Nandinder K.;
APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
APPLICANT: LEE, Sally; BECHA, Shanya D.;
APPLICANT: LEE, Soo Yeun; TRAN, Dyen K.;
APPLICANT: ELLIOTT, Vicki S.; LUO, Wen;
APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
APPLICANT: LU, Yan; ZEBARADIAN, Yeganeh
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PF-1040 USN
CURRENT APPLICATION NUMBER: US/10/480,988
CURRENT FILING DATE: 2003-12-16
PRIOR APPLICATION NUMBER: PCT/US02/19360
PRIOR FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/300,508
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: US 60/303,445
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/305,405
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/311,442
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 60/314,821
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/315,992
PRIOR FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: US 60/378,205
PRIOR FILING DATE: 2002-05-03
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PERL Program
SEQ ID NO 8
LENGTH: 414
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7632424CD1
US-10-480-988-8

```

Query Match	100.0%	Score 2125;	DB 30;	Length 414;
Best Local Similarity	100.0%	Pred. No. 1.2e-214;		
Matches 414;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0
Qy	1	MLILTKTGVGFPEKPKRYEFLRSEFNPHPGTLPLHKILVLGIETSCDDTPAAAVDEGNV	60	
Db	1	MLILTKTGVGFPEKPKRYEFLRSEFNPHPGTLPLHKILVLGIETSCDDTPAAAVDEGNV	60	
Qy	61	LGEAHSQTEVHLKTGGIYPPAAQQLHRENIQRIVOBALISASGVSPPDLSAIAITTKPGL	120	
Db	61	LGEAHSQTEVHLKTGGIYPPAAQQLHRENIQRIVOBALISASGVSPPDLSAIAITTKPGL	120	
Qy	121	ALSLGVGVSFSQILVGQLKKPPIPIHHMBANALTRLTNKVPFPLVLLISGCHCLALV	180	
Db	121	ALSLGVGVSFSQILVGQLKKPPIPIHHMBANALTRLTNKVPFPLVLLISGCHCLALV	180	
Qy	181	QGVSPDFLLGKSLIDYAPGDMLDKVARBSLSLHGPECSMSGKALBHLAKQGRPHPIIX	240	
Db	181	QGVSPDFLLGKSLIDYAPGDMLDKVARBSLSLHGPECSMSGKALBHLAKQGRPHPIIX	240	
Qy	241	PPLHAKNCDSPFTGILQHYTIDKILKKKKSEBIEIKGQILISSAADYAAVQHTMACHLYR	300	
Db	241	PPLHAKNCDSPFTGILQHYTIDKILKKKKSEBIEIKGQILISSAADYAAVQHTMACHLYR	300	
Qy	301	THRAILFCKQBDLIFONNAVLVASGVASNFYIRBALEILTNAQCCTLLCPBPLCTDNG	360	

```

Db          301  THRLLPCKQKDLIPONNAVLVSGVSNFYIRALAEILTNATCTLLCPPLCTDNG 360
QY          361  IMIANGIERLRAGIGILHDIEGIRYEPKCPGLVDISKEVGBASIKVPLKMEI 414
Db          361  IMIANGIERLRAGIGILHDIEGIRYEPKCPGLVDISKEVGBASIKVPLKMEI 414

RESULT 4
US-10-649-273-2
: Sequence 2, Application US/10649273
: GENERAL INFORMATION:
: APPLICANT: Bristol-Myers Squibb Company
: TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
: FILE REFERENCE: D0073 CNT
: CURRENT APPLICATION NUMBER: US/10/649,273
: CURRENT FILING DATE: 2003-08-27
: PRIOR APPLICATION NUMBER: US 60/266,518
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 10/067,443
: PRIOR FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US 60/282,814
: PRIOR FILING DATE: 2001-04-10
: NUMBER OF SEQ ID NOS: 71
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 2
: LENGTH: 414
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-649-273-2

```

	Query Match	100.0%;	Score 2125;	DB 32;	Length 414;	
	Best Local Similarity	100.0%;	Pred. No. 1,2e-214;			
	Matches 414;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MLITLTAGVFPFKSKRYEPLERSFNHFGTLLFLHKTVLGIETSCDDTTAAAVDETGVN	60			
Dy	1	MLITLTAGVGFPEPSKRYVEPERSFNFHPGTLFLHKTVLGIEISCDDTAAAVDETONV	60			
QY	61	LGEAHSOTEVHLTKGTGIVPPAQOOLRENIQRIVQELSLASGSVPSSLAIATIKPGL	120			
Dy	61	LGEAHSOTEVHLTKGTGIVPPAQOOLRENIQRIVQELSLASGSVPSSALAITIKGL	120			
QY	121	ALSIVGLSFSLOLVQLKKPFIDIHMEAHALTIRLTNKVEPPFLVLLISGGHCLLAV	180			
Dy	121	ALSIVGLSFSLSLOLVQLKKPFIDIHMEAHALTIRLTNKVEPPFLVLLISGGHCLLAV	180			
QY	181	QGVSDFLLTKKSULIAPGDMDKYARRLSLIKHEPCSTMSGKALEHLAKGNRHFPIK	240			
Dy	181	QGVSDFLLTKKSULIAPGDMDKYARRSLIKHEPCSTMSGKALEHLAKGNRRHFPIK	240			
QY	241	PLLHAANCDSPFTGLQHTVDKIIMKEKEBEGIKGOQLISSAADIAATVOHTMACLYKR	300			
Dy	241	PLLHAANCDSPFTGLQHTVDKIIMKEKEBEGIKGOQLISSAADIAATVOHTMACLYKR	300			
QY	301	THRALIFCKRDILLPONNAVVAAGVASNYIRRAEILTNACOTLLCPPRCLCTNG	360			
Dy	301	THRALIFCKRDILLPONNAVVAAGVASNYIRRAEILTNACOTLLCPPRCLCTNG	360			
QY	361	IMIAMNGIEBLRAGLIGILHDIEGIRYEKCPLGYDISKEVEBASIKVQLMDEI	414			
Dy	361	IMIAMNGIEBLRAGLIGILHDIEGIRYEKCPLGYDISKEVEBASIKVQLMDEI	414			
 RESULT 5 US-10-651-722-2 ; Sequence 2, Application us/10651722 ; GENERAL INFORMATION: ; APPLICANT: Biotech Weapons Squibb Company ; TITLE OR INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1 ; FILE REFERENCE: D0073 DIV ; CURRENT APPLICATION NUMBER: US/10/651,722 ; CURRENT FILING DATE: 2003-08-29						

; PRIOR APPLICATION NUMBER: US 60/266,518
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: US 10/067,443
 ; PRIOR FILING DATE: 2002-02-05
 ; PRIOR APPLICATION NUMBER: US 60/282,814
 ; PRIOR FILING DATE: 2001-04-10
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 2
 ; LENGTH: 414
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; US-10-651-722-2

Query Match 100.0%; Score 2125; DB 32; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFPKSKRKYVEFLRSFNHFGTLFLHKIVLGIETSCDDTAAAVDDEGNV 60
 DB 1 MLILTKTAGVFFPKSKRKYVEFLRSFNHFGTLFLHKIVLGIETSCDDTAAAVDDEGNV 60
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPDLSAIAATTIKPGL 120
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPDLSAIAATTIKPGL 120
 QY 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEPPLVLLISGHCLLALV 180
 DB 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEPPLVLLISGHCLLALV 180
 QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSITMSGKAIEMHAKQGRFFPDIK 240
 DB 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSITMSGKAIEMHAKQGRFFPDIK 240
 QY 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMACHLVKR 300
 DB 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMACHLVKR 300
 QY 301 THRALLFCQKQDRLPQNNNAVIVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQKQDRLPQNNNAVIVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360
 QY 361 IMIANNGERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414
 DB 361 IMIANNGERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

RESULT 6

US-60-266-518-2

; Sequence 2, Application US/60266518
 ; GENERAL INFORMATION:
 ; APPLICANT: Bristol-Myers Squibb Company
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
 ; TITLE OF INVENTION: SPINAL CORD, MP-1
 ; FILE REFERENCE: D0073 PSP
 ; CURRENT APPLICATION NUMBER: US/60/266,518
 ; CURRENT FILING DATE: 2001-02-05
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 414
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; US-60-266-518-2

Query Match 100.0%; Score 2125; DB 37; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFPKSKRKYVEFLRSFNHFGTLFLHKIVLGIETSCDDTAAAVDDEGNV 60
 DB 1 MLILTKTAGVFFPKSKRKYVEFLRSFNHFGTLFLHKIVLGIETSCDDTAAAVDDEGNV 60

QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPDLSAIAATTIKPGL 120
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPDLSAIAATTIKPGL 120
 QY 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEPPLVLLISGHCLLALV 180
 DB 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEPPLVLLISGHCLLALV 180
 QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSITMSGKAIEMHAKQGRFFPDIK 240
 DB 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSITMSGKAIEMHAKQGRFFPDIK 240
 QY 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMACHLVKR 300
 DB 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMACHLVKR 300
 QY 301 THRALLFCQKQDRLPQNNNAVIVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQKQDRLPQNNNAVIVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360
 QY 361 IMIANNGERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414
 DB 361 IMIANNGERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

RESULT 7

US-60-282-814-2

; Sequence 2, Application US/60282814
 ; GENERAL INFORMATION:
 ; APPLICANT: Bristol-Myers Squibb Company
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
 ; TITLE OF INVENTION: SPINAL CORD, MP-1
 ; FILE REFERENCE: D0073 PSP1
 ; CURRENT APPLICATION NUMBER: US/60/282,814
 ; CURRENT FILING DATE: 2001-04-10
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 414
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; US-60-282-814-2

Query Match 100.0%; Score 2125; DB 37; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFPKSKRKYVEFLRSFNHFGTLFLHKIVLGIETSCDDTAAAVDDEGNV 60
 DB 1 MLILTKTAGVFFPKSKRKYVEFLRSFNHFGTLFLHKIVLGIETSCDDTAAAVDDEGNV 60
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPDLSAIAATTIKPGL 120
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPDLSAIAATTIKPGL 120
 QY 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEPPLVLLISGHCLLALV 180
 DB 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEPPLVLLISGHCLLALV 180
 QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSITMSGKAIEMHAKQGRFFPDIK 240
 DB 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSITMSGKAIEMHAKQGRFFPDIK 240
 QY 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMACHLVKR 300
 DB 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMACHLVKR 300
 QY 301 THRALLFCQKQDRLPQNNNAVIVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQKQDRLPQNNNAVIVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360
 QY 361 IMIANNGERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414
 DB 361 IMIANNGERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

Db 361 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 414

RESULT 8

US-10-918-754-1372

Sequence 1372, Application US/10918754

GENERAL INFORMATION:

APPLICANT: CARCILL, Michele

TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND

TITLE OF INVENTION: USES THEREOF

FILE REFERENCE: C0001480

CURRENT APPLICATION NUMBER: US/10/918,754

CURRENT FILING DATE: 2004-08-16

NUMBER OF SEQ ID NOS: 91238

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1372

LENGTH: 425

TYPE: PRT

ORGANISM: Homo sapiens

US-10-918-754-1372

Query Match 100.0%; Score 2125; DB 35; Length 425;

Best Local Similarity 100.0%; Pred. No. 1.3e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MLILTKTAGVFFPKSKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVVDETGNY 60

12 MLILTKTAGVFFPKSKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVVDETGNY 71

61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAATITIKPGL 120

72 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAATITIKPGL 131

121 ALSIGVLSFSIQLVGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180

132 ALSIGVLSFSIQLVGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 191

181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKQGRFHPDIK 240

192 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKQGRFHPDIK 251

241 PPLHAKNCDFSTGLQHTVDKIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300

252 PPLHAKNCDFSTGLQHTVDKIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 311

301 THRALLFCQKQDILLPONNAVIVASGVASNFYIRRALEILTNATQCTLLCPPRLCTDNG 360

312 THRALLFCQKQDILLPONNAVIVASGVASNFYIRRALEILTNATQCTLLCPPRLCTDNG 371

361 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 414

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

Query Match 100.0%; Score 2090.5; DB 1; Length 439;

Best Local Similarity 93.8%; Pred. No. 6e-211;

Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

1 MLILTKTAGVFFPKSKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVVDETGNY 60

12 MLILTKTAGVFFPKSKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVVDETGNY 71

61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAATITIKPGL 120

72 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAATITIKPGL 131

121 ALSIGVLSFSIQLVGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180

132 ALSIGVLSFSIQLVGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 191

181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKQGRFHPDIK 240

192 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKQGRFHPDIK 251

241 PPLHAKNCDFSTGLQHTVDKIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300

252 PPLHAKNCDFSTGLQHTVDKIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 311

301 THRALLFCQKQDILLPONNAVIVASGVASNFYIRRALEILTNATQCTLLCPPRLCTDNG 360

312 THRALLFCQKQDILLPONNAVIVASGVASNFYIRRALEILTNATQCTLLCPPRLCTDNG 371

361 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 414

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKVEGASIKVPOLKMEI 425

```

Db      181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKAIEHLAKQGNRFPDIX 240
QY      241 PPLHAKNCDFSPFTGLQHTVDKIKMKKEKEBGI-----EK 275
Db      241 PPLHAKNCDFSPFTGLQHTVDKIKMKKEKEBGIPLISVEQINIPGLCLKIAHFCRYEK 300
QY      276 GQILSSADIAATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 335
Db      301 GQILSSADIAATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 360
QY      336 ALBITLTAOCTLLCPPRCLCTDNGIMIAMNGIERLRAGLIGLHDIGIRYBPKCPLGYD 395
Db      361 ALBITLTAOCTLLCPPRCLCTDNGIMIAMNGIERLRAGLIGLHDIGIRYBPKCPLGYD 420
QY      396 ISKEVGEASIKVPOLKMEI 414
Db      421 ISKEVGEASIKVPOLKMEI 439

RESULT 11
US-10-649-273-19
; Sequence 19, Application US/10649273
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

Query Match      98.4%; Score 2090.5; DB 32; Length 439;
Best Local Similarity 93.8%; Pred. No. 6e-211;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY      1 MLILTKTAGVFPKSKRKVYEFELRSFNFHFGTLFLHKIVLGIETSCDDTPAAVVDETGNY 60
Db      1 MLILTKTAGVFPKSKRKVYEFELRSFNFHFGTLFLHKIVLGIETSCDDTPAAVVDETGNY 60
QY      61 LGEAHSOTEVHLKTGQIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAIAATTIKPGL 120
Db      61 LGEAHSOTEVHLKTGQIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAIAATTIKPGL 120
QY      121 ALSLVGSLFSFSLQVQLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
Db      121 ALSLVGSLFSFSLQVQLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
QY      121 ALSLVGSLFSFSLQVQLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
Db      121 ALSLVGSLFSFSLQVQLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
QY      181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKAIEHLAKQGNRFPDIX 240
Db      181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKAIEHLAKQGNRFPDIX 240
QY      241 PPLHAKNCDFSPFTGLQHTVDKIKMKKEKEBGI-----EK 275
Db      241 PPLHAKNCDFSPFTGLQHTVDKIKMKKEKEBGIPLISVEQINIPGLCLKIAHFCRYEK 300
QY      276 GQILSSADIAATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 335
Db      301 GQILSSADIAATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 360
QY      336 ALBITLTAOCTLLCPPRCLCTDNGIMIAMNGIERLRAGLIGLHDIGIRYBPKCPLGYD 395
Db      361 ALBITLTAOCTLLCPPRCLCTDNGIMIAMNGIERLRAGLIGLHDIGIRYBPKCPLGYD 420

```

```

QY      396 ISKEVGEASIKVPOLKMEI 414
Db      421 ISKEVGEASIKVPOLKMEI 439

RESULT 12
US-10-651-722-19
; Sequence 19, Application US/10651722
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

Query Match      98.4%; Score 2090.5; DB 32; Length 439;
Best Local Similarity 93.8%; Pred. No. 6e-211;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY      1 MLILTKTAGVFPKSKRKVYEFELRSFNFHFGTLFLHKIVLGIETSCDDTPAAVVDETGNY 60
Db      1 MLILTKTAGVFPKSKRKVYEFELRSFNFHFGTLFLHKIVLGIETSCDDTPAAVVDETGNY 60
QY      61 LGEAHSOTEVHLKTGQIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAIAATTIKPGL 120
Db      61 LGEAHSOTEVHLKTGQIVPPAAQOLHRENIQRIYOEALASAGVSPDLSAIAATTIKPGL 120
QY      121 ALSLVGSLFSFSLQVQLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
Db      121 ALSLVGSLFSFSLQVQLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
QY      181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKAIEHLAKQGNRFPDIX 240
Db      181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKAIEHLAKQGNRFPDIX 240
QY      241 PPLHAKNCDFSPFTGLQHTVDKIKMKKEKEBGI-----EK 275
Db      241 PPLHAKNCDFSPFTGLQHTVDKIKMKKEKEBGIPLISVEQINIPGLCLKIAHFCRYEK 300
QY      276 GQILSSADIAATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 335
Db      301 GQILSSADIAATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 360
QY      336 ALBITLTAOCTLLCPPRCLCTDNGIMIAMNGIERLRAGLIGLHDIGIRYBPKCPLGYD 395
Db      361 ALBITLTAOCTLLCPPRCLCTDNGIMIAMNGIERLRAGLIGLHDIGIRYBPKCPLGYD 420
QY      396 ISKEVGEASIKVPOLKMEI 414
Db      421 ISKEVGEASIKVPOLKMEI 439

RESULT 13
US-60-266-518-19
; Sequence 19, Application US/60266518
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1

```

FILE REFERENCE: D0073 PSP
 CURRENT APPLICATION NUMBER: US/60/266,518
 CURRENT FILING DATE: 2001-02-05
 NUMBER OF SEQ ID NOS: 28
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 19
 LENGTH: 439
 TYPE: PRT
 ORGANISM: homo sapiens
 US-60-266-518-19

Query Match 98.4%; Score 2090.5; DB 37; Length 439;
 Best Local Similarity 93.8%; Pred. No. 6e-211;
 Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFPKSRKRYVEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60
 DB 1 MLITKTAGVFPKSRKRYVEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIVOEALSASGVSPDLSAIAITTKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIVOEALSASGVSPDLSAIAITTKPGL 120
 QY 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCCLALV 180
 DB 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCCLALV 180
 QY 181 QGVSDFLLGKSLDIPAGMDLKVARRSLIKPEECSTMSGGKAIEHLAKQGNRFHFDIK 240
 DB 181 QGVSDFLLGKSLDIPAGMDLKVARRSLIKPEECSTMSGGKAIEHLAKQGNRFHFDIK 240
 QY 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGI-----EK 275
 DB 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGI-----EK 275
 QY 276 GQILSSADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 335
 DB 301 GQILSSADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 360
 QY 336 AEIITNATQCTLLCPPLCTDNGIMIAMNGIERLPAGLGIHDIIEGIRYBPCPLGVD 395
 DB 361 AEIITNATQCTLLCPPLCTDNGIMIAMNGIERLPAGLGIHDIIEGIRYBPCPLGVD 420
 QY 396 ISKEVGEASIKVPLQKMEI 414
 DB 421 ISKEVGEASIKVPLQKMEI 439

RESULT 14
 US-60-282-814-19

Sequence 19, Application US/60282814
 GENERAL INFORMATION:
 APPLICANT: Bristol-Myers Squibb Company
 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
 TITLE OF INVENTION: SPINAL CORD, MP-1
 FILE REFERENCE: D0073 PSP
 CURRENT APPLICATION NUMBER: US/60/282,814
 CURRENT FILING DATE: 2001-04-10
 NUMBER OF SEQ ID NOS: 32
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 19
 LENGTH: 439
 TYPE: PRT
 ORGANISM: homo sapiens
 US-60-282-814-19

Query Match 98.4%; Score 2090.5; DB 37; Length 439;
 Best Local Similarity 93.8%; Pred. No. 6e-211;
 Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFPKSRKRYVEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60
 DB 1 MLITKTAGVFPKSRKRYVEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60

QY 61 LGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIVOEALSASGVSPDLSAIAITTKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIVOEALSASGVSPDLSAIAITTKPGL 120
 QY 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCCLALV 180
 DB 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCCLALV 180
 QY 181 QGVSDFLLGKSLDIPAGMDLKVARRSLIKPEECSTMSGGKAIEHLAKQGNRFHFDIK 240
 DB 181 QGVSDFLLGKSLDIPAGMDLKVARRSLIKPEECSTMSGGKAIEHLAKQGNRFHFDIK 240
 QY 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGI-----EK 275
 DB 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGI-----EK 275
 QY 276 GQILSSADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 335
 DB 301 GQILSSADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 360
 QY 336 AEIITNATQCTLLCPPLCTDNGIMIAMNGIERLPAGLGIHDIIEGIRYBPCPLGVD 395
 DB 361 AEIITNATQCTLLCPPLCTDNGIMIAMNGIERLPAGLGIHDIIEGIRYBPCPLGVD 420
 QY 396 ISKEVGEASIKVPLQKMEI 414
 DB 421 ISKEVGEASIKVPLQKMEI 439

RESULT 15
 US-10-012-140-5

Sequence 5, Application US/10012140
 GENERAL INFORMATION:
 APPLICANT: Leiby, Kevin R.
 APPLICANT: Kapeller-Libermann, Rosana
 TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
 TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
 FILE REFERENCE: 381552004900
 CURRENT APPLICATION NUMBER: US/10/012,140
 CURRENT FILING DATE: 2001-11-08
 PRIOR APPLICATION NUMBER: 60/246,768
 PRIOR FILING DATE: 2000-11-08
 PRIOR APPLICATION NUMBER: 60/246,772
 PRIOR FILING DATE: 2000-11-08
 PRIOR APPLICATION NUMBER: 60/249,185
 PRIOR FILING DATE: 2000-11-15
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5
 LENGTH: 414
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-012-140-5

Query Match 98.3%; Score 2088; DB 26; Length 414;
 Best Local Similarity 98.3%; Pred. No. 1e-210;
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLITKTAGVFPKSRKRYVEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60
 DB 1 MLITKTAGVFPKSRKRYVEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIVOEALSASGVSPDLSAIAITTKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIVOEALSASGVSPDLSAIAITTKPGL 120
 QY 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCCLALV 180
 DB 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCCLALV 180

```

Qy 181 QGVSDPILLGKSLDIAPGMDLVARRSLIKHPECSTMSGKAIENLAKQNRFPDIK 240
Db 181 QGVSDPILLGKSLDIAPGMDLVARRSLIKHPECSTMSGKAIENLAKQNRFPDIK 240
Qy 241 PPLHAKNCDFSPYGLQHTDXTIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR 300
Db 241 PPLHAKNCDFSPYGLQHTDXTIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR 300
Qy 301 THRALFCOKORDLPONNAVIVASGVASNFYIRALFILTMATQCTLLCPPRLCTDNG 360
Db 301 THRALFCOKORDLPONNAVIVASGVASNFYIRALFILTMATQCTLLCPPRLCTDNG 360
Qy 361 IMIAMNGIERLRAGIGILHDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 414
Db 361 IMIAMNGIERLRAGIGILHDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 414

```

Search completed: February 16, 2005, 13:18:55
 Job time : 229.8 secs


```

Db      178 -TYPAGRIBDLAHQGD-I-YDPPAMIKEDNLEFSPGLKSAFNLHNAE-----QKG 230
QY      277 QILSSADIAATVQHTMACHLVKTRTHAILFKQKRDLLPQNNNAVVASGVASNFYIRRA 336
Db      231 ESIST-BDLKASFOAAVMDILMAKTKKAL-----EYYPVK--TLVAGGVANAKGLRER 281
QY      337 LEILTNATQCTLLCPPRCTDNGIMIA-----NN 366
Db      282 L--AAETDVKVITIPRLCGDNAGMTAYASVSEWN 315

```

RESULT 2

```

US-11-027-843-4857
; Sequence 4857, Application US/11027843
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangong
; APPLICANT: Opperman, Timothy
; APPLICANT: Housewear, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: 3687.1000-024
; CURRENT FILING DATE: US/11/027,843
; PRIOR APPLICATION NUMBER: US 10/640,833
; PRIOR FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4857
; LENGTH: 336
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
US-11-027-843-4857

```

Query Match 21.3%; Score 453; DB 7; Length 336;

Best Local Similarity 34.2%; Pred. No. 1.2e-33;

Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;

```

QY      37 KIVLGIEISCDPTAAVVDETGNVIGEAIHSOTVEHLKTGGIVPPAAOQLHRENIQRIYQ 96
Db      4 RYILAFETSCDETSVAVLKNDDELISNVIASQIESHKRRGGVPEVASRHHVEVITACIE 63
QY      97 EALSASGVSPDLSAATTIKRGALSLGVGLSFLQVLGOLKRPPIPIHNEAHLTIR 156
Db      64 EALAEGITEBDVTAAVATYVGPGLVGLVGLSAKAFAMAGLPLIPNHAAGHLMMAQ 123
QY      157 LTNKVEPFVLLISGCHLALVQGVSDPFLIGSLDIAPGMDLKVARRSLIKHPEC 216
Db      124 SVEPLEPFLALLVSGHTELVYSEADYKIVGETRDVAVEADVKGRVWGL----- 177
QY      217 STMSGKALEHIAKQGNRFHDIKPRLIHAKNCDSPFTGLQHTVDKIIMKEKEBGIKKG 276
Db      178 -TYPAGRIBDLAHQGD-I-YDPPAMIKEDNLEFSPGLKSAFNLHNAE-----QKG 230
QY      277 QILSSADIAATVQHTMACHLVKTRTHAILFKQKRDLLPQNNNAVVASGVASNFYIRRA 336
Db      231 ESIST-BDLKASFOAAVMDILMAKTKKAL-----EYYPVK--TLVAGGVANAKGLRER 281
QY      337 LEILTNATQCTLLCPPRCTDNGIMIA-----NN 366
Db      282 L--AAETDVKVITIPRLCGDNAGMTAYASVSEWN 315

```

RESULT 3

```

US-11-027-878-4857
; Sequence 4857, Application US/11027878
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangong
; APPLICANT: Opperman, Timothy
; APPLICANT: Housewear, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: 3687.1000-008
; CURRENT FILING DATE: US/11/027,878
; CURRENT FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: US 10/640,833
; PRIOR FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4857
; LENGTH: 336
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
US-11-027-878-4857

```

Query Match 21.3%; Score 453; DB 7; Length 336;

Best Local Similarity 34.2%; Pred. No. 1.2e-33;

Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;

```

QY      37 KIVLGIEISCDPTAAVVDETGNVIGEAIHSOTVEHLKTGGIVPPAAOQLHRENIQRIYQ 96
Db      4 RYILAFETSCDETSVAVLKNDDELISNVIASQIESHKRRGGVPEVASRHHVEVITACIE 63
QY      97 EALSASGVSPDLSAATTIKRGALSLGVGLSFLQVLGOLKRPPIPIHNEAHLTIR 156
Db      64 EALAEGITEBDVTAAVATYVGPGLVGLVGLSAKAFAMAGLPLIPNHAAGHLMMAQ 123
QY      157 LTNKVEPFVLLISGCHLALVQGVSDPFLIGSLDIAPGMDLKVARRSLIKHPEC 216
Db      124 SVEPLEPFLALLVSGHTELVYSEADYKIVGETRDVAVEADVKGRVWGL----- 177
QY      217 STMSGKALEHIAKQGNRFHDIKPRLIHAKNCDSPFTGLQHTVDKIIMKEKEBGIKKG 276
Db      178 -TYPAGRIBDLAHQGD-I-YDPPAMIKEDNLEFSPGLKSAFNLHNAE-----QKG 230
QY      277 QILSSADIAATVQHTMACHLVKTRTHAILFKQKRDLLPQNNNAVVASGVASNFYIRRA 336
Db      231 ESIST-BDLKASFOAAVMDILMAKTKKAL-----EYYPVK--TLVAGGVANAKGLRER 281
QY      337 LEILTNATQCTLLCPPRCTDNGIMIA-----NN 366
Db      282 L--AAETDVKVITIPRLCGDNAGMTAYASVSEWN 315

```

RESULT 4

```

US-11-028-169-4857
; Sequence 4857, Application US/11028169
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangong
; APPLICANT: Opperman, Timothy
; APPLICANT: Housewear, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: 3687.1000-010
; CURRENT FILING DATE: US/11/028,169

```


Db 64 BALBAGITBEDVAVAVTYPGVLGALLVGLSAKAPAMAGLPLIPVNHMAHGLMAAQ 123
QY 157 LTNKVEPFLVLLISGCHCLALVGVSDPFLLSGLDAPGMDLKVARRSLIKHPEC 216
Db 124 SVEPLEPFLALVSGHTELTVYVSEAGDYKIVGSTRDAVGEAYDKVGRVWGL----- 177
QY 217 STMGGKAIEMHAKOGRFHFDPKPLHAKNCDSFTGLOHVTDKTIKKKEEGIBKG 276
Db 178 -TYPAGRIDELAHQODI-YDPPRAMIKEDNLEFSFGLSKAFINLHNAL-----QKG 230
QY 277 QILSSADIAATVOHTMACLVKTRTHAILFCQBDLLPONNAVIVASGVASNFYIRRA 336
Db 231 ESLST-EDLCASFOAVMDIMAKTKKAL-----BEYVVK--TLVAGGVANKGLRER 281
QY 337 LEILTNATQCTLLCPPLCTDNGIMIA-----WN 366
Db 282 L--AAITDVKVIIPRLCGDNAGMIAYASVSEWN 315

RESULT 7

US-11-027-879-4857
; Sequence 4857, Application US/11027879
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qidong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: 3687,1000-021
; CURRENT APPLICATION NUMBER: US/11/027,879
; PRIOR FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: US 10/640,833
; PRIOR FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4857
; LENGTH: 336
; TYPE: PRP
; ORGANISM: Streptococcus pneumoniae
US-11-027-879-4857

Query Match 21.3%; Score 453; DB 7; Length 336;
Best Local Similarity 34.2%; Pred. No. 1.2e-33;
Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;

QY 37 KIVLGIETSCDPTAAVVDGNTVGEAIIHSGTEVHLKGTGIVPPAAOOLHRENIORIVQ 96
Db 4 RYILAFETSCDPTSAVAVLKNDELLSNVIAQSIESHKRGVPEVASRHVETIACIE 63
QY 97 EALSASGVSPDLSAIIATIKPGLALSLGVGLSFSIQLVGOLKXPPPIPIHMEAHALTR 156
Db 64 BALBAGITBEDVAVAVTYPGVLGALLVGLSAKAPAMAGLPLIPVNHMAHGLMAAQ 123
QY 157 LTNKVEPFLVLLISGCHCLALVGVSDPFLLSGLDAPGMDLKVARRSLIKHPEC 216
Db 124 SVEPLEPFLALVSGHTELTVYVSEAGDYKIVGSTRDAVGEAYDKVGRVWGL----- 177
QY 217 STMGGKAIEMHAKOGRFHFDPKPLHAKNCDSFTGLOHVTDKTIKKKEEGIBKG 276
Db 178 -TYPAGRIDELAHQODI-YDPPRAMIKEDNLEFSFGLSKAFINLHNAL-----QKG 230
QY 277 QILSSADIAATVOHTMACLVKTRTHAILFCQBDLLPONNAVIVASGVASNFYIRRA 336
Db 282 L--AAITDVKVIIPRLCGDNAGMIAYASVSEWN 315

Db 231 ESLST-EDLCASFOAVMDIMAKTKKAL-----BEYVVK--TLVAGGVANKGLRER 281
QY 337 LEILTNATQCTLLCPPLCTDNGIMIA-----WN 366
Db 282 L--AAITDVKVIIPRLCGDNAGMIAYASVSEWN 315

RESULT 8

US-11-028-149-4857
; Sequence 4857, Application US/11028149
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qidong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: 3687,1000-017
; CURRENT APPLICATION NUMBER: US/11/028,149
; PRIOR FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: US 10/640,833
; PRIOR FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4857
; LENGTH: 336
; TYPE: PRP
; ORGANISM: Streptococcus pneumoniae
US-11-028-149-4857

Query Match 21.3%; Score 453; DB 7; Length 336;
Best Local Similarity 34.2%; Pred. No. 1.2e-33;
Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;

QY 37 KIVLGIETSCDPTAAVVDGNTVGEAIIHSGTEVHLKGTGIVPPAAOOLHRENIORIVQ 96
Db 4 RYILAFETSCDPTSAVAVLKNDELLSNVIAQSIESHKRGVPEVASRHVETIACIE 63
QY 97 EALSASGVSPDLSAIIATIKPGLALSLGVGLSFSIQLVGOLKXPPPIPIHMEAHALTR 156
Db 64 BALBAGITBEDVAVAVTYPGVLGALLVGLSAKAPAMAGLPLIPVNHMAHGLMAAQ 123
QY 157 LTNKVEPFLVLLISGCHCLALVGVSDPFLLSGLDAPGMDLKVARRSLIKHPEC 216
Db 124 SVEPLEPFLALVSGHTELTVYVSEAGDYKIVGSTRDAVGEAYDKVGRVWGL----- 177
QY 217 STMGGKAIEMHAKOGRFHFDPKPLHAKNCDSFTGLOHVTDKTIKKKEEGIBKG 276
Db 178 -TYPAGRIDELAHQODI-YDPPRAMIKEDNLEFSFGLSKAFINLHNAL-----QKG 230
QY 277 QILSSADIAATVOHTMACLVKTRTHAILFCQBDLLPONNAVIVASGVASNFYIRRA 336
Db 231 ESLST-EDLCASFOAVMDIMAKTKKAL-----BEYVVK--TLVAGGVANKGLRER 281
QY 337 LEILTNATQCTLLCPPLCTDNGIMIA-----WN 366
Db 282 L--AAITDVKVIIPRLCGDNAGMIAYASVSEWN 315

RESULT 9

US-11-027-802-4857
; Sequence 4857, Application US/11027802
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David

RESULT 14

US-11-027-844-4857

; Sequence 4857, Application US/11027844

; GENERAL INFORMATION:

; APPLICANT: Doucette-Stamm, Lynn

; APPLICANT: Bush, David

; APPLICANT: Zeng, Qilandong

; APPLICANT: Opperman, Timothy

; APPLICANT: Housewart, Chad Eric

; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

; FILE REFERENCE: 3687.1000-012

; CURRENT APPLICATION NUMBER: US/11/027,844

; PRIORITY FILING DATE: 2004-12-30

; PRIORITY FILING DATE: 2003-08-14

; PRIORITY FILING DATE: 2000-05-26

; PRIORITY FILING DATE: 1998-06-30

; PRIORITY FILING DATE: 1998-05-12

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; FILE REFERENCE: 3687.1000-023

; CURRENT APPLICATION NUMBER: US/11/028,050

; CURRENT FILING DATE: 2004-12-30

; PRIORITY FILING DATE: 2004-12-30

; PRIORITY FILING DATE: 2003-08-14

; PRIORITY FILING DATE: 2000-05-26

; PRIORITY FILING DATE: 1998-06-30

; PRIORITY FILING DATE: 1998-05-12

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

; PRIORITY FILING DATE: 1997-07-02

Query Match 21.3%; Score 453; DB 7; Length 336;

Best Local Similarity 34.2%; Pred. No. 1.2e-33;

Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;

US-11-027-844-4857

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

; TYPE: PR

; LENGTH: 336

; ORGANISM: Streptococcus pneumoniae

Search completed: February 16, 2005, 13:19:29
Job time : 15.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:10:32 ; Search time 6588 Seconds
(without alignments)
3045.000 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 2125
Sequence: 1 MLILITKAGVFPKRSKRKY.....DISKEVGASIKVQLKMEI 414

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 segs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xjh
-Q=/cgn2_1/USPTO.spool/US10649273/runat_14022005_114702_16399/app.query.fasta_1.1429
-DB=GenEmbl -OPMT=faetap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=50 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blowum62 -TRANS=human40.cdi -LIST=45
-DOCALLGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPM=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CGN 1.1 8655 @runat_14022005_114702_16399 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	1908	9	BC011904 Homo sapi
2	2125	100.0	2197	6	AR428803 Sequence
3	2090.5	98.4	1387	6	AR428808 Sequence
4	2090.5	98.4	1387	9	HS4295148 Homo sapi

5	2088	98.3	1245	6	AX664697 Sequence
6	2088	98.3	1820	6	AX664695 Sequence
7	1944	91.5	2208	6	AX713716 Sequence
8	1944	91.5	2208	9	AK055441 Homo sapi
9	1835	86.4	1844	10	BC058172 Homo sapi
10	1747	82.2	1416	6	AR541929 Sequence
11	1725	81.2	1546	10	BC078974 Rattus no
12	1385	65.2	1566	6	AR428809 Sequence
13	1362	64.1	1522	5	BX934991 Gallus ga
14	1273	59.9	1017	10	BC038910 Gallus ga
15	1208	56.8	1558	5	BX930963 Gallus ga
16	1193.5	56.2	8415	9	AC013468 Homo sapi
17	1186.5	55.8	14364	6	AR428807 Sequence
18	995.5	46.8	249601	2	AC114153 Rattus no
19	995.5	46.8	308652	2	AC121478 Rattus no
20	970	45.6	256751	10	AC122925 Mus muscu
21	950.5	44.7	1109	10	BC051211 Mus muscu
22	938	44.1	860	5	BX930694 Gallus ga
23	827	38.9	875	6	CQ721898 Sequence
24	751.5	35.4	121251	5	AL591593 Zebrafish
25	714.5	33.6	1576	3	AY051882 Drosophi
26	714.5	33.6	1601	6	CQ606432 Sequence
27	710	33.4	1474	3	AK113378 Clona int
28	709	33.4	117322	5	AL672217 Zebrafish
29	700.5	33.0	1385	6	BD157102 Primer fo
30	700.5	33.0	1385	6	AX878239 Sequence
31	700.5	32.4	1385	9	AK027836 Homo sapi
32	687.5	32.4	3656	6	CQ606431 Sequence
33	687.5	32.4	14679	2	AC018262 Drosophi
34	687.5	32.4	180263	3	AC010671 Drosophi
35	687.5	32.4	207432	3	AE003513 Drosophi
36	668.5	31.5	1443	8	AY024338 Arabidops
37	668.5	31.5	1474	8	AY117283 Arabidops
38	668.5	31.5	1567	8	AY063864 Arabidops
39	662.5	31.2	1557	8	AY084577 Arabidops
40	656	30.9	1672	8	AK070912 Oryza sat
41	629	29.6	571	6	BD154975 Primer fo
42	629	29.6	571	6	AX874913 Sequence
43	618	29.1	110000	2	AP006501_08 Continuation (9 of
44	577.5	27.2	260271	1	AE017258 Wolbachia
45	568.5	26.8	333800	1	SM591792 Sinorhizo

ALIGNMENTS

RESULT 1	BC011904	1908 bp	mRNA	linear	PRI 23-DEC-2003
LOCUS	BC011904	Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA			
DEFINITION	clone MGC:20293 IMAGE:4121450), complete cds.				
ACCESSION	BC011904.2	GI:40225818			
VERSION	BC011904				
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 1908) Straussberg, R.L., Feingold, R.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shennan, C.M., Schler, G.D., Altshul, S.P., Zeeberg, B., Buetow, K.H., Scheffer, C.P., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, Y., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Scheetz, T.B., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carrinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Kullany, S.J., Bosak, S.A., McEwan, P.D., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wooley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, B., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,				

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.B., Shterf, A., Schein, J.B., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 12477932

2 (bases 1 to 1908)

Strasberg, R. Direct Submission Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NH-MGC Project URL: <http://mgc.nci.nih.gov>

On Dec 19, 2003 this sequence version replaced gi:15080281.

Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Issue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hgrl.nih.gov
Ahner, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Masker, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W., Teursson, C., Vogt, J.L., Walker, M.A., Weherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILML at: <http://image.llnl.gov>
Series: IRAL Plate: 28 Row: 1 Column: 22.
Location/Qualifiers

FEATURES

SOURCE

gene

CDS

misc_feature

ORIGIN

Alignment Scores:

Pred. No.: 6,176-171 Length: 1908
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x BC011904 (1-1908)

1 MetLeuIleLeuThrThyThrAlaGlyValPhePhePheProSerTybArgIysValTyr 20
104 AGCTAACTTACCTAAGCTCAGAGAGTTTAAACCAATCAAAAGGAAGTTTAT 163
21 GluPheLeuArgSerPheAspPheHisProGlyTyrLeuPheLeuHisIleValLeu 40
164 GAATTTTAAAGAGTTTATTTATTTCTCTGAGACACTTTCTTCAATAATAGATTG 223
41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
224 GGAATTTAAAGAGTTTATTTATTTCTCTGAGACACTTTCTTCAATAATAGATTG 283
61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleuLeuThrGlyIleValPro 80
284 TTGGAGAGAGCAATACATTCCTCAAACTGAAGTTCAATTTAAACAGGTGGATTGCT 343
81 ProAlaAlaGlnGluLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
344 CAGAGAGCTCAAGAGTTTACAGAGAAATATTTCAAGAAATGACAGAAAGCTTTTCT 403
101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleuPheProGlyLeu 120
404 GCCATGTAGAGTCTCTCCAGTACCTCTGAGCAATTTGCAACATTAACACAGACTT 463
121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyLeuLeuIlys 140
464 GCTTTAAGCTGGAGTGGCTTATCATTTAGCTTACACTGATGAGACAGTTAAAG 523
141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160
524 CATTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 583
161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
584 GTCGATTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 643
181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMet 200
644 CAAGAGTTTCAGATTTTCTGCTTCTGGAAGCTTTGGAATAGCACAGGTGACATG 703
201 LeuAspIysValAlaArgArgLeuSerLeuIleIysHisProGluCysSerThrMetSer 220
704 CTTCAGCAAGGTGCAAGAGACTTTCTTTAAATTAACATCCAGAGCTCCACATGAGT 763
221 GlyGlyIysAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHisAspAspIleIys 240
764 GGTGGGAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAA 823
241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
824 CTTCCCTTCATCAGCTAAATATTTGATTTTCTTTTACGACCTTCAACACGTTACT 883
261 AspIysIleIleMetIysIysGluIysGluIysGlyIleGluIysGlyIleLeuSer 280
884 GATTAATTAATATGAAAAAGAAAAAGGAAGGATTTGAGAAAGGCAAAATCCGTCT 943
281 SerAlaAlaAspIleAlaAlaThrValGlnHisSerMetAlaCysHisIleuValIysArg 300
944 TCAGCAGCAGACATTCCTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1003
301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAlaIleVal 320
1004 ACAATCGGCTATTCTGTTTGTAAAGCAGAGAGCTTTTACTCTCAAAATATGACAGTA 1063

QY 321 LeuValAlaSerGlyGlyValAlaSerAspNheTyrIleArgArgAlaLeuGluIleLeu 340
DB 1064 CTGGTTCACCTGGTGGTGGCAAGTAATTCTATATCCGACAGGCTCTGGAAATTTTAA 1123
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProAlaGluCysThrAspAsnGly 360
DB 1124 ACAACGCAACACAGTGCACCTTGTGTGTCTCTCCACACATATGACATGATATATGGC 1183
QY 361 IleMetIleAlaTPrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1184 ATTATGATTCATGAGAAATGCGATTGAAAGACTACGTGCTGGCTGGCAATTTTACATGAC 1243
QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerIysGlyVal 400
DB 1244 ATGAAAGGCATCGCTATGAAACCAAAATGTCCTCTTGAGTAAATATCAATCAAAAGACTT 1303
QY 401 GlyGluAlaSerIleLysValProGluLeuLysMetGluIle 414
DB 1304 GGAGAAAGCTTCATMAAAGTACCACAATTAATAATGAGATA 1345

RESULT 2
AR428803 AR428803 2197 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6642041.
ACCESSION AR428803
VERSION AR428803.1 GI:40186589
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2197)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;
FEATURES
source 1. 2197
location/Qualifiers
source /organism="Unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 7.37e-171 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x AR428803 (1-2197)

QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGluLeuValGlyIleLeuLys 140
DB 591 GCTTTAAGCTTGGAGTGGGCTTATCATTTAGCTTACACTGGTACGACGTTAAAAAAG 650
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 651 CCATTCAATCCCATTCATCATATGAGGCTCATGCACTTACTATTTAGTTGATGACCAATAAA 710
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyHisGlyLeuLeuAlaLeuVal 180
DB 711 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACTGTCTGTTGGCATTAAGT 770
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 771 CAGAGATTCAGATTTTCTGCTTCTTGAAAAGCTCTTGACATACACCAAGTGCATG 830
QY 201 LeuAspLysValAlaArgArgLeuSerIleLysHisProGluCysSerThrMetSer 220
DB 831 CTGACAAAGTGGCAAGAAAGACTTCTTAAATAAAATCCAGAGTCTCCACCATGAGT 890
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 891 GGTGGAAAAGCCATGAAACATTTGGCCAAACAGAAATGAAATTTGATTTTGCATCAAAA 950
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 951 CTTCCCTTCATCATAGCTTAAATAATGTAATTTTCTTTTACTGGACTTCACACGTTACT 1010
QY 261 AspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280
DB 1011 GATTAATAATATATGAAAAAGAAAAAGGAAGATTTGAGAAGGGGCAATCCGTGCT 1070
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB 1071 TCAGCAGCAGACATTTGCTGCCACAGTACAGCACACATGCACTGTCATCTTGGAANA 1130
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
DB 1131 ACACATCGGGCTAATTCGTGTTTGTAAAGCAGAGAAATTTGTACTTCAATAATATGACATG 1190
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340
DB 1191 CTGGTTCATCTGGTGGTGTCCCAAGTAATCTTATATCCGACAGACTCTGAAAAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
DB 1251 ACAACGCAACACAGTGCACCTTGTGTCTCTCCACAGACTATGACATGATAATGGC 1310
QY 361 IleMetIleAlaTPrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1311 ATTATGATTCATGGAATATGATTTGAAAGACATACGCTGGCTGGCAATTTTACATGAC 1370
QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerIysGlyVal 400
DB 1371 ATGAAAGCATTCGCTATGAAACCAAAATGTCCTCTTGAGTAAATATCAAAAGAAATT 1430
QY 401 GlyGluAlaSerIleLysValProGluLeuLysMetGluIle 414
DB 1431 GGAGAAAGCTTCATMAAAGTACCACAATTAATAATGAGATA 1472

RESULT 3
AR428808 AR428808 1387 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 21 from patent US 6642041.
ACCESSION AR428808
VERSION AR428808.1 GI:40186594
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1387)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.

TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;
FEATURES Location/Qualifiers
source 1..1387
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 3.54e-168 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: 6 Gaps: 1

US-10-649-273-2 (1-414) x AF428808 (1-1387)

QY 1 MetLeuIleLeuThrIleThyThraGlyValIlePhePheIysProSerIysArgIysValIleTyr 20
DB 24 ATGCTAATCTTGACTAAGACACGACGAGTTTATTAACATCAAAAGAAAGTTTAT 83
QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisIleValIleu 40
DB 84 GAATTTTAAAGAAAGTTTAAATTTTCATCTGAAACATATTCTTCATTAATATGATTG 143
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValIleAspGluThrGlyAsnVal 60
DB 144 GGAATTGAATTAAGTGTGATGATACACAGCTGCTGCGTGAGTAAACGTGAATATG 203
QY 61 LeuGlyGluAlaIleHisSerGluThrGlyValHisLeuIysThrGlyIleValPro 80
DB 204 TTGGGAAGAGCAATACATTTCCCAACCTGAAGTTCACTTAATAACAGGTGGATTGTCT 263
QY 81 ProAlaIleGluGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 264 CCACAGCTCAACAGCTTCACAGAAATATTCACAGAAATATTCACAGAAAGCTTTCT 323
QY 101 AlaSerGlyValSerProSerThrAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
DB 324 GCCAGTGAAGCTCTCCCAAGTGAACCTTCACGAATTCGAACATCAACACAGACTT 383
QY 121 AlaLeuSerLeuGlyValIleGlyLeuSerPheSerLeuGlnLeuValIleGlyLeuIys 140
DB 384 GCTTTAAGCTGGAGAGTGGCTTATCATTTACCTACAGCTGGTGAACAGTTAAAG 443
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160
DB 444 CCATTCAATCCATTCAATCATATGAGGCTCATGCACTTACATTAAGTTGACCAATAA 503
QY 161 ValGluPheProPheLeuValIleLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 504 GTAATAATTTCTTTTATGTTCTTTTATGTTCTTGAGAGTCACTGCTGTGGCATTAAGT 563
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMet 200
DB 564 CAAGAGATTTCAGAAATTTCTGCTTCTGAAAGCTTTTGAAATAGCCACAGTAACTAG 623
QY 201 LeuAspIysValAlaArgArgLeuSerLeuIleIysHisProGluCysSerThrMetSer 220
DB 624 CTTGACAAAGGTGGCAAGAGACTTTCTTTAATTAACATCCAGAGTGTCCACCATGAGT 683
QY 221 GlyGlyIysAlaIleGlnHisLeuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240
DB 684 GGTGGAAAGCCATAGAGCATTTGGCCAAACAGAAATAGATTTCATTGTGACATCAAA 743
QY 241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 744 CCTCCCTTGCAATGCTTAAATTTGTATTTTCTTTTACGTGACCTCAACACGTTACT 803
QY 261 AspIysIleIleMetIleValGlyGluIysGluGluGlyIle----- 273
DB 804 GATTAATATTAATGAAGAAAGAAAGAAAGAGTATATTTCTTAATTAAGTAAAGTTGAA 863

QY 274 -----GluIys 275
DB 864 CAGATAAATATTCCTGATGTCGCTTAATAAGCTCTCATTTTCGACGATAGGAAG 923
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295
DB 924 GGGCAATCTGCTTCACAGACAGCATTTGCTGCACAGTACAGCAACAATGCGATGT 983
QY 296 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuPro 315
DB 984 CATCTTGAAAAGAACATCGCGGTATTCGTTTGTGAAGACAGACATTTGTACTT 1043
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspMetThrIleArgArg 335
DB 1044 CAAATTAATGCAATCTGCTGCTCATCTGAGTGCAGAGTCACTTCATATCCCAAG 1103
QY 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355
DB 1104 GCTTGAAATTTTAAACAAGCAACAGTGCACCTTTGTGTCTCTCCCAAGACTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375
DB 1164 TGCACGTAAATGCGATTAATGATTCATGCAATGGAATGTAAGAACTACGTGCTTG 1223
QY 376 GlyIleLeuHisAspIleGluGlyIleArgGlyTrpGluProIysCysProLeuGlyValAsp 395
DB 1224 GGCATTTTACATGACATGAAAGCAATCCGCTTAAGAACAAATGTCCTCTGAGTAGAC 1283
QY 396 IleSerIysGluValIleGlyAlaSerIleIysValProGlnLeuIysMetGluIle 414
DB 1284 AATTCAAAGAGTGGAGAACCTTCCTAATAAGTACCAATTAATAATGAGATA 1340
RESULT 4
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000
LOCUS Homo sapiens mRNA for putative sialoglycoprotease type 2.
DEFINITION
ACCESSION AJ295148
VERSION AJ295148.1 GI:11071726
KEYWORDS metalloproteinase; sialoglycoprotease.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 Chen, J.M., Fortunato, M. and Barrett, A.J.
Cloning and sequencing of a second human putative
sialoglycoprotease homologue
Unpublished
2 (bases 1 to 1387)
Chen, J.M.
Direct Submission
Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,
UNITED KINGDOM
FEATURES
source
1..1387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/feature_type="uterns"
24..1343
/function="metalloproteinase of family M22"
/codon_start=1
/product="putative sialoglycoprotease type 2"
/protein_id="CAC14666.1"
/db_xref="GI:11071727"
/db_xref="GOA:Q9H4B0"
/db_xref="UniProt/TREMBL:Q9H4B0"
/translation="MULTITAGVFFKSKRYEFLRSPNHPETLFLAKIVLGIET
SCDDRAAVVDETVNLGEALISQTEVHLKGTGIVPAACQQLRENIQIVBALAS
CVSPSDLAIAATTIKRGALSLGVGPSLOLVGOLKRFPIIHMEARLITRLNK
VEFPPLVLILSGHCLIALVQSVDFLLGKSLDIAPGMDLKVARRSLIKIPBST

MSGGKALIEHLAKQGNRFHFDIKPRLHAKNCDPSFTGLQHVTDKIMKKEBEGIFLI
SKYBOINI PGLICTIAIAHFCRYEKGOILSSADIAATVOHTMAKCLVTRTHAILFK
ORDILPQNNANVLVSGGVAASFYIRRALEIETNATOCITLCPPLCTDNGIMLWNG
IERLRGILGILHDEGIRYERKCPGLVDISKEVBASIKVQLKMEI*

ORIGIN

Alignment Scores:
Pred. No.: 3,54e-168 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatch: 2
Query Match: 98.38% Indels: 25
Gaps: 1

US-10-649-273-2 (1-414) x HSA295148 (1-1387)

QY 1 MetLeuIleLeuThrLySerThrAlaGlyValPhePheLyProSerLyAspGlyValTyr 20
DB 24 ATGCTAACTTGAAGTAAAGTCAAGAGAGTCTTTTAAACCATCAAAAGGAAAGCTTAT 83
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLeuValLeu 40
DB 84 GAATTTTAAAGAGTTTAAATTTTATTCCTGAAACACTATTTCTTCATTAATAATAGTATTG 143
QY 41 GlyIleGlyThrSerCyAspAspThrAlaAlaValValAspGlyThrGlyAsnVal 60
DB 144 GGAATTAAGTAACTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 203
QY 61 LeuGlyValAlaIleHisSerGlyThrGlyValHisLeuLySerGlyValIleValPro 80
DB 204 TTGGAGAAAGCAATATACATTTCCCAAGTGAAGTCAATTTAAACAGGTGGATGTTCTT 263
QY 81 ProAlaAlaGlnGlnLeuHisAspGlyLeuAsnIleGlnAlaGlyValGlnLeuVal 100
DB 264 CCAGCAGCTCAACAGCTTCAAGAGAAATATTCACCAAAATAGTCAAGAAAGCTTTCT 323
QY 101 AlSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleAspProGlyLeu 120
DB 324 GCCAGTGAAGTCTTCCAGTGAAGTCTTCCAGCAATTCAGCAATTAACCAAGCAAGCTT 383
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleLeuVal 140
DB 384 GCTTTAAGCCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGATGAGACAGTTAAAG 443
QY 141 ProPheIleProIleHisSerGlyAlaHisAlaLeuThrIleArgLeuThrAsnVal 160
DB 444 CCAATTCATTCCTCAATTCATCATATGAGGCTCATGCACTTACTATGAGTTCGCAATA 503
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyHisCysLeuLeuAlaLeuVal 180
DB 504 GTAGAAATTTCTTTTATGATTTCTTGAATTCGAGGTCACTGCTGTGCAATTAGTT 563
QY 181 GlnGlyValSerAspPheLeuLeuGlyLySerSerLeuAspIleAlaProGlyAspMet 200
DB 564 CAAGAGATTTCAAGATTTTCTGCTTCTGGAAGCTTTTGACATAGCACAGTACATG 623
QY 201 LeuAspLyValAlaArgArgLeuSerLeuIleLySerHisProGlnCysSerThrMetSer 220
DB 624 CTTCACAGAGGTGGAGAGAGCTTTCTTAAATTAACATCCAGAGTGTCCACCATAGT 683
QY 221 GlyGlyLyValAlaIleGlyHisLeuAlaLyGlnGlyAsnArgPheHisPheAspIleVal 240
DB 684 GTGGGAAACCCATAGAGCATTTGGCCAAACAGAAATAGATTTCAATTTGACATCAAA 743
QY 241 ProProLeuHisHisAlaLyAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 744 CTTCCCTTGCACTACGCTAAATATGATTTTCTTTTACTGACCTTCACACACCTTACT 803
QY 261 AspLyIleIleMetLySlyGlyLySlyGlyLySlyLeu----- 273
DB 804 GATTAATTAATTAATGAAGAAAGAGAGAGATTAATTTCTTAATTAAGTAAAGTGA 863
QY 274 -----GluLys 275

DB 864 CAGATAATATATTCCTGATTTGCTCAAAAATAGCTGCATTTCTGACGATAGAGAG 923
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295
DB 924 GAGCAATATCCTGCTTCCAGACAGACATTCCTGCAAGTACAGACACAAATGCAATGT 983
QY 296 HisLeuValLyAspArgThrHisArgAlaIleLeuPheCysLySerGlnArgAspLeuPro 315
DB 984 CATCTTGTAAGAAAGACATCGGCTATTCCTTTGTAAGACAGAGACTTGTTACT 1043
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheYrIleArgArg 335
DB 1044 CAAATTAATGACGATCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1103
QY 336 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlyLeu 355
DB 1104 GCTCTGGAATTTTAAACCAACCAACAGTGCATTTGTTGTCCTCTCCAGACTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 375
DB 1164 TGCATGATTAATGCTATTAATGATGATGATGATGATGATGATGATGATGATGATG 1223
QY 376 GlyIleLeuHisAspIleGlyValIleArgGlyGlyProLySerCysProLeuGlyValAsp 395
DB 1224 GGCATTTTATCATGATAGAGAGATCCGCTATGAAACCAAAATGTCCTTGAGATAGAC 1283
QY 396 HisSerLySlyGlyValGlyValAlaSerIleLyValProGlnLeuLySerGlyLeu 414
DB 1284 ATATCAAAAGAGTGGAGAGCTTCCATTAATAATGATCAAAATTAATAATGAGATA 1340

RESULT 5
AX664697 1245 bp DNA linear PAT 22-MAR-2003
LOCUS AX664697 Sequence 6 from Patent WO02074960.
DEFINITION AX664697
ACCESSION AX664697.1 GI:29164457
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Leiby,K.R., Kapeller-Libermann,R. and Gluckmann,M.
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
of human proteins and uses thereof
JOURNAL Patent: WO 02074960-A 6 26-SEP-2002;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source 1..1245
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 5.03e-168 Length: 1245
Score: 2088.00 Matches: 407
Percent Similarity: 99.03% Conservative: 3
Best Local Similarity: 98.31% Mismatches: 4
Query Match: 98.26% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x AX664697 (1-1245)
QY 1 MetLeuIleLeuThrLySerThrAlaGlyValPhePheLyProSerLyAspGlyValTyr 20
DB 1 ATGCTAACTTGAAGTAAAGTCAAGAGAGTCTTTTAAACCATCAAAAGGAAAGCTTAT 60
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLeuValLeu 40
DB 61 GAATTTTAAAGAGTTTAAATTTTATTCCTGAAACACTATTTCTTCATTAATAATAGTATTG 120

```

Qy 41 GlyIleGIuThrSerCyAspAspThrAlaAlaValAspGIuThrGIyAsnVal 60
Db 121 GGAATTGAACTAGTTGATGATACACAGCTGCTGCTGATGAACTGGAATGTG 180
Qy 61 LeuGIyGIuAlaIleHisSerGIuThrGIuValHisLeuLySerThyGIyGIyIleValPro 80
Db 181 TTGGAGAGGAACAATACATTTCCCAACTGAAAGTTCATTTAAACAAGTGGGATGTTCTCT 240
Qy 81 ProAlaAlaGIuThrLeuHisArgGIuAsnIleGIuArgIleValGIuGIuAlaLeuSer 100
Db 241 CCAGAGCTCAACAGCTTCACAGAAATATTCACCAATAGTACCAAGAAAGCTCTTCT 300
Qy 101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleYSerProGIyLeu 120
Db 301 GCCAGTGAAGTCTCTCCAGAGTACCTCTCAGCAATTCGACATACCAATAAACAGAGACTT 360
Qy 121 AlaLeuSerLeuGIyValGIyLeuSerPheSerLeuGIuLeuValGIyGIuLeuLyS 140
Db 361 GCTTTAAGCTGGGAGTGGGCTTATCATTTAGCTTACAGCTGGTACAGCACTTAAAG 420
Qy 141 ProPheIleProIleHisIleMetGIuAlaHisAlaIleuThrIleArgLeuThrAsn 160
Db 421 CCATTCAATCCCATTCATCATATGAGAGGCTCATGCACTTACTATTAGTTGACCAATAA 480
Qy 161 ValGIuPheProPheLeuValLeuLeuIleSerGIyGIyHisCyLeuLeuAlaLeuVal 180
Db 481 GTAGAAATTTCTTTTATGTTCTTTTGAATTTCTGAGAGTCACTGCTGTGGCATTTAGT 540
Qy 181 GIuGIyValSerAspPheLeuLeuLeuGIyLySerLeuAspIleAlaProGIyAspMet 200
Db 541 CAAGAGATTCAGATTTTCTGCTCTTGAAAGCTTTTGACATGACACAGTGAACATG 600
Qy 201 LeuAspLyValAlaArgArgLeuSerIleuIleYHisIleProGIuCySerThrMetSer 220
Db 601 CTTCACAAAGGCGGCAAGAGACTTCTTAAATAAACAATCCAGAGTCCCACTAGAGT 660
Qy 221 GIyGIyValAlaIleGIuHisLeuAlaLySGInGIyAsnArgPheHisPheAspIleY 240
Db 661 GGTGGAAAGCCATAGAACATTTGGCCCAAGCAAGAAATAGATTTTATTTGACATTCAA 720
Qy 241 ProProLeuHisHisAlaLyAsnCyAspPheSerPheThrGIyLeuGIuHisValThr 260
Db 721 CCTCCCTTCGACATCAGTAAATTTGTGATTTTCTTTTACGACCTTCAACAGCTTACT 780
Qy 261 AspLySerIleIleMetLyLyGIuLyGIuGIyGIyIleGIuLyGIyGIuIleLeuSer 280
Db 781 GATTAATAATTAATGAACACAGAAACAAAGAGGATATGAGAAAGGCGCAATCCCTGCT 840
Qy 281 SerAlaAlaAspIleAlaIleThrValGIuHisIleThrMetAlaCyHisIleuValLyArg 300
Db 841 TCAGACAGACATTTGCTGCCAGGTACACACAAATGGCAATGTCATCTTGTGAAGAA 900
Qy 301 ThrHisArgAlaIleLeuPheCyLySGInArgAspLeuLeuProGIuHisAsnAlaVal 320
Db 901 ACAATCGGGCTATTCCTTTTGTATGACAGAGAGACTTGTATCTCAAAATATATGACAGTA 960
Qy 321 LeuValAlaSerGIyGIyValAlaSerAsnPheYIleArgArgAlaLeuGIuIleLeu 340
Db 961 CTGGTTCATCTGGGCTGGTGGCAAGTAATCTTATCCGACAGCTCTGCAAAATTTTAA 1020
Qy 341 ThrAsnAlaThrGIuCySerThrLeuLeuCySerProArgArgLeuCySerThrAspAsn 360
Db 1021 ACAAAACGACACAGTGCATTTGTGTCTCTCCACAGCTATGCACTATTAATGAGC 1080
Qy 361 IleMetIleAlaIleAspAsnGIyIleGIuArgLeuArgAlaGIyLeuGIyIleLeuHisAsp 380
Db 1081 ATTATGATTCATGCAATGATGATTAACAAAGACTACGCTGCTGGCATTTTATCATGAC 1140
Qy 381 IleGIuGIyIleArgYIleArgIuProLyCySerProLeuGIyValAspIleSerLyGIuVal 400
Db 1141 ATAGAAAGGCAATCCGCTATGAACCAAAATGTCTTGGAGTGAATATCAAAAGAGTT 1200
Qy 401 GIyGIuAlaSerIleYValProGIuLeuYSerMetGIuIle 414

```

```

Db 1201 GGAGAGCTTCATTAAGTACCAACATTAATAAGAGATATA 1242
RESULT 6
AX664695
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.
38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
of human proteins and uses thereof
Patent: WO 02074960-A 4 26-SEP-2002;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..1820
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
146..1390
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD80044.1"
/db_xref="GI:29164456"
/translation="MULTITAGVFPKSKRYEFLRSPNFPGTLPLKIVLGIST
SCDDTAAAVVDSTGVIGEAHISQTHVHLKTGGIYPPAPOOLHRETIORIVQALAS
GYSBPDLSAIRATTIRPGALSLSGVLSRLQVGGKRPFIHHHEAALTRITLNK
VEFPVLVLTISGHCILALVGVSPFLKSLSDLPGLMDKVRARLSLIRHPEST
MSGRAIETALVQGNRPHFDIPLPHANCPSTGLQHVTDKNNENKQSEIGK
QILSSAADIATVQHTMACHLVETRRAILPKQRDLFPONNAVVAAGSVASNFYIR
RLRLITMNAQCTLCPPRLCTDNGIMANNGIRLRAGLGIHDIIEGIRYEPKCP
GVDSIKYGEASIKYPOLKMET"
ORIGIN
Alignment Scores:
Pred. No.: 8,14e-168 Length: 1820
Score: 2088.00 Matches: 407
Percent Similarity: 99.03% Conservative: 3
Best Local Similarity: 98.31% Mismatches: 4
Query Match: 98.26% Indels: 0
DB: Gaps: 0
US-10-649-273-2 (1-414) x AX664695 (1-1820)
Qy 1 MetLeuIleLeuThrIleYThrAlaGIyValPhePheLyProSerIleYArgLyValIYr 20
Db 146 ATGCTAATCTTCACTAAGACTCAGAGAGTATTTTAAACATCAAAAGAAAGTTAT 205
Qy 21 GIuPheLeuArgSerPheAsnPheHisIleProGIyThrLeuPheLeuHisIleValIleu 40
Db 206 GAATTTTAAAGATTTTAATTTTCACTCTGGAACCTATTTCTTATTAATAATAGTATG 265
Qy 41 GlyIleGIuThrSerCyAspAspThrAlaAlaValAspGIuThrGIyAsnVal 60
Db 266 GGAATTGAACTAGTTGATGATACACAGCTGCTGTGGATGAATGAATGGAATGTG 325
Qy 61 LeuGIyGIuAlaIleHisSerGIuThrGIuValHisLeuLySerThyGIyGIyIleValPro 80
Db 326 TTGGAGAGACATACATTTCCCAACTGAAGTTCAATTTAAACAAGTGGGATGTTCTCT 385
Qy 81 ProAlaAlaGIuThrLeuHisArgGIuAsnIleGIuArgIleValGIuGIuAlaLeuSer 100
Db 386 CCAGAGCTCAACAGCTTCACAGAAATATTCACCAAGATGTACACAGAAAGCTCTTCT 445
Qy 101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleYSerProGIyLeu 120
Db 446 GCCAGTGAAGTCTCTCCAGTACCTCTCAAGCAATTCGACATTCATTAACACAGACTT 505

```

QY 121 AAlaSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuValys 140
DB 506 GCTTAAAGCTGGAGGCTGCTTATCATTTAGCTACAGCTGGTAGACAGTAAAAAG 565
QY 141 ProPheLeuProIleHisIleMetCysAlaHisAlaLeuThrIleArgLeuThrAnlys 160
DB 566 CCATTTCATTCCTCATTCATCATATGAGGCTCATGCACTTACTATTAGGTTGACCAATAAA 625
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyValHisCysLeuLeuAlaLeuVal 180
DB 626 GTAAATTTCTTTTATTTAGTTTGTGATTTCTGAGAGGCTGCTGTGTGACATTAGTT 685
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 686 CAAGAGATTGAGATTTTCTGCTTCTTGAAAGCTTTTGACATACACAGTGCATG 745
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
DB 746 CTGACCAAGGTGGCAAGAAAGCTTTCTTAATAAACATCCAGAGTCTCCACATGAGT 805
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAspArgPheHisPheAspIleLys 240
DB 806 GGTGGGAAAGCCATPAGAACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAA 865
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 866 CTTCCCTTGCACTATGCTTAAATTTGTATTTTCTTTTACCTGACCTTCAACACCTTACT 925
QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGluLysGlyGlnIleLeuSer 280
DB 926 GATTAATAATATAGAAAACAGAAACAGAAAGGATTTGAGAAAGGGCAATTCGTCT 985
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisSerMetAlaCysHisLeuValLysArg 300
DB 986 TCAGCACAGACATGCTGCCACAGTACAGACACAAATGCAATGCTATCTGTGAAAAAG 1045
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnLysArgAspLeuProGlnAsnAsnAlaVal 320
DB 1046 ACACATGGGCTATTCGTGTTTGTAAGCAAGAGCTGTGTAACCTCAAAATATATCAGTA 1105
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
DB 1106 CTGTTTCATCTGCTGTGTGTGCAAGTACTTCTATATCCGAGAGCTCTGGAATTTTA 1165
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgGluCysThrAspAsnGly 360
DB 1166 ACAAAACCAACACAGTGCACTTGTGTGTCTCTCCACAGATATGCACTGATATGAGC 1225
QY 361 IleMetIleAlaTyrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1226 ATATATGATTCATGAAATGTAATGAAAGACTACCTGTGCTGGGATTTTATATATAC 1285
QY 381 IleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlyVal 400
DB 1286 ATAGAAGGCATCCCTATGAACCAAAATGCTCTTGAAGTAGACATATCAAAAGAGTT 1345
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1346 GGAAGAGCTTCATTAATAAGTACCAATTAATAATAATGAGATTA 1387

RESULT 7
AX713716 2208 bp DNA linear PAT 15-APR-2003
LOCUS Sequence 400 from Patent EP1293569.
DEFINITION AX713716
ACCESSION AX713716
VERSION AX713716.1 GI:29888642
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1

AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
Masuko, Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 400 19-MAR-2003;
Helix Research Institute (Jp) ; Research Association for
Biotechnology (Jp)
FEATURES
source location/Qualifiers
1..2208 /organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 1,83e-155 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 6 Gaps: 1
US-10-649-273-2 (1-414) x AX713716 (1-2208)
QY 1 MetLeuIleLeuThrIleThrAlaGlyValAlaPhePheLysProSerLysArgLysValTyr 20
DB 344 ATGCTATATCTTGACTAGACGCTGCAAGAGTTTATTTTAAACATCAAAAGAAAGTTAT 403
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
DB 404 GATTTTAAAGAAAGTTTATTTATTTCTGGAACCTATTTCTTCAATAAATATGATTTG 463
QY 41 GlyIleGlnThrSerCysAspAspThrAlaAlaValAlaLysGlnLysArgVal 60
DB 464 GGAATTAAGAACTAGTTGTATGATATACAGCAGCTGTGTGTGATGAAACTGGAATATGTG 523
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro 80
DB 524 TTGGGAGAAAGCAATATATTCCTCAAACTGAAGTTCAATTAATAACAGGTGGATTTGCTT 583
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 584 CAGAGCTCAACAGCTTCAACAGAAATATTCAACGAATATCAAGAGCTCTTTCT 643
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrIleLysProGlyLeu 120
DB 644 GCCAGTGGAGTCTCTCCAAAGTGACTCTCAGCAATTCGCAATTAACCAATTAACAGAGCTT 703
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
DB 704 GCTTTAAAGCTGGAGTGGCTTATATTAAGTTTACAGTGTGTAGACAGTTAAAAAG 763
QY 141 ProPheIleProIleHisIleMetCysAlaHisAlaLeuThrIleArgLeuThrAnlys 160
DB 764 CCATTTCATTCCTCATTCATCATATGAGGCTCATGCACTTACTATTAGGTTGACCAATAAA 823
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyValHisCysLeuLeuAlaLeuVal 180
DB 824 GTAAATTTCTTTTATTTAGTTTGTGATTTCTGAGAGTCACTGTCTGTTGGATTAAGTT 883
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 884 CAAGGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGACATACACCAAGGTGACATG 943
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
DB 944 CTGACCAAGGTGGCAAGAAAGCTTTCTTAATAAACATCCAGAGTCTCCACAGATGACT 1003
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAspArgPheHisPheAspIleLys 240
DB 1004 GGTGGGAAAGCCATPAGAACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAA 1063

QY 241 ProProLeuHsiHsiSaIaIyAsnCyAspPheSerPheThrGlyLeuGlnHsiValThr 260
 Db 1064 CctCCCTTGATCATGCTAAATAATTTGATTTCTTTTACGACCTTCAACACCTTACT 1123
 QY 261 AspLysIleIleMetLysLysGlnLysGlnGluGlyIleGluLysGlyGlnIleLeuSer 280
 Db 1124 GATAAATTAATAAATGAAAAAGAAAAAGGAAAGGATATGAGAAAGGCAATCTGCTCT 1183
 QY 281 SerIlaIaIaPpIleAlaIaIaThrValGlnHsiThrMetAlaCyHsiIleValLysArg 300
 Db 1184 TCGACGACGACATGCTGCGCACGATACGACACAAAGGCGATGCTCTGTGAAAAAGA 1243
 QY 301 ThrHsiArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal 320
 Db 1244 ACACATCGGGGTATCTCTTTTGTAGACAGAGACTTCTTACCTCAAAATTAATGACGTA 1303
 QY 321 LeuValAlaSerGlyGlyValAlaIaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeu 340
 Db 1304 CTGGTTCATCTGGTGGTGGTGGCAAGTAATCTGTATCCGACAGCTCTGGAAATTTTAA 1363
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCySerProProProArgGluCysThrAspAsnGly 360
 Db 1364 ACAACACCAACACAGTGCATCTTGTGTCTCTCCACAGCTATGACATGATATATGAC 1423
 QY 361 IleMetIleAlaTPaenGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHsiAsp 380
 Db 1424 ATTTATGATGCA----- 1435
 QY 381 IleGluGlyIleArgTyrgluProLysCysProLeuGlyValAspIleSerLysGlyVal 400
 Db 1436 -----TGATGTCCTCTTGAGATGACATATCAAAAGAGATT 1471
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
 Db 1472 GGAGAGACCTTCATTAATAAGTACCAATTAATAAATGAGAGATA 1513
 RESULT 8
 LOCUS AK055441 2208 bp mRNA linear PRI 30-JAN-2004
 DEFINITION Homo sapiens cDNA FLJ30879 f1s, clone FEBRA2004592, highly similar
 to Homo sapiens mRNA for putative sialoglycoprotease type 2.
 ACCESSION AK055441
 VERSION AK055441.1 GI:16550166
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS 1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
 Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
 Sekine, M., Ohtsuka, M., Nishi, T., Shibahara, T., Tanaka, T.,
 Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
 Nagahori, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
 Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
 Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
 Kikkawa, B., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
 Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
 Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
 Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
 Yosiida, M., Hottura, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
 Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
 Takeuchi, K., Arita, M., Imose, N., Muesashi, K., Yuuki, H., Oshima, A.,
 Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
 Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
 Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizuguchi, H.,
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
 Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
 Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
 Tashiro, H., Tanigami, A., Fujitawa, T., Ono, T., Yamada, K., Fujii, Y.,
 Otsaki, K., Hira, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
 Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
 Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
 Oyama, M., Hara, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
 Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
 Nagase, T., Nomura, N., Kikuchi, H., Maehno, Y., Yamaishi, R.,
 Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
 Complete sequencing and characterization of 21,243 full-length
 human cDNAs
 Nat. Genet. 36 (1), 40-45 (2004)
 14702039
 2
 Nishi, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H.,
 Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
 Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
 Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
 Kikuchi, K., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,
 Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
 Sugano, S., Nagahori, K., Maehno, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 3 (bases 1 to 2208)
 REFERENCE
 AUTHORS 3 Isogai, T., Otsuki, T. and Sugiyama, T.
 TITLE Direct Submission
 JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: RAB and
 HRI.
 FEATURES
 source 1..2208
 Location/Qualifiers
 1..2208
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="FEBRA2004592"
 /tissue_type="brain"
 /clone_1b="FEBRA2"
 /dev_stage="fetus"
 /note="cloning vector: pME189FL3"
 344..1438
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB70923.1"
 /db_xref="GI:16550167"
 /translation="MLILTKTAGVPFKSKRVYERLRSFNPFGTLPLHKIVLGIET
 SCDDTRAAVNDGTGAVLGEALHSQTBVHLKTGCIYPPAAQQLHRENTQIVQELASAS
 GVSFSDALAIATTTPGLALSLGVLSFSLQVGLQKRPFIPIHNEALATIRLTNK
 VEPFLVLLISGCHRLALVQGVDFLLKGLSLDAPGMDLRVARRPLIRKPEST
 MSGGAIHLIAKCPFHFDIKPPLHAKNCDPFGQGVHVDKILMKKEKEGIEKG
 OILSSAIIIAAVVQHTMACHVYRTHRALIFCKQBDLPLQNNNAVLAASGVASNFICR
 RALELTITATGCTCLCPPRLCTDNGIMIA"
 CDS
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,83e-155 Length: 2208
 Score: 1944.00 Matches: 366
 Percent Similarity: 93.24% Conservative: 0
 Best Local Similarity: 93.24% Mismatches: 4
 Query Match: 91.48% Indels: 24
 DB: 9 Gaps: 1
 US-10-649-273-2 (1-414) x AK055441 (1-2208)
 QY 1 MetLeuIleLeuThryThraIaGlyValPhePheLysProSerLysArgLysValIlyr 20
 Db 344 ARGCTAAATCTTGACATGACAGCTCAGAGAGTTTATTAACCATCAAAAGAAAGTTTAT 403
 QY 21 GlupheLeuArgSerPheasnPheHsiPProGlyThrLeuPheLeuHsiLysIleValLeu 40

Db 404 GAATTTTAAAGAGTTTAAATTTTCATCTCGAGACATATTTCTCATAAAAATGATTTG 463
 Qy 41 G1Y1leGIuThrSerCyAspAspThrAlaAlaValAlaAspGIuThrGIYAsnVal 60
 Db 464 GGAATTTGAAATCTAGTGTGATGATACACAGAGCTGTGTGTGATGATGAAATGTG 523
 Qy 61 LeuGIYGIuAlaIleHisSerGIuThrGIuValHisLeuYleThrGIYGIYIleValPro 80
 Db 524 TTGGAGAGAGCAATACATTTCCCAAACTGAAGTTCAATTTAAAAACAGTGGATTTCTCT 583
 Qy 81 ProAlaIleAGInGInLeuHisArgGIuAsnIleGIuValIleValGIuGIuAlaLeuSer 100
 Db 584 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATGATCAGAAAGCTCTTCT 643
 Qy 101 AlaSerGIYAlaSerProSerAspLeuSerAlaIleAlaThrThrIleYAspProGIYLeu 120
 Db 644 GCCAGTGAAGTCTCTCCAGAGTACCTCTCAGCAATTCGACATTAACATTAACAGAGCTT 703
 Qy 121 AlaLeuSerLeuGIYAlaGIYLeuSerPheSerLeuGIuValGIYGIuLeuYAsnYs 140
 Db 704 GCTTTAAGCCTGGAGTGGGCTTATCATTTAGCTTACAGCTGTGAGACAGTTAAAAAG 763
 Qy 141 ProPheIleProIleHisIleMetGIuAlaHisAlaLeuThrIleArgLeuThrAsnYs 160
 Db 764 CCATTCATTTCCATTCATCATATGAGAGGCTCATATGCACTTACTATTAGTTGACCAATAA 823
 Qy 161 ValGIuTherProPheLeuValLeuLeuIleSerGIYGIYHisCysLeuLeuAlaLeuVal 180
 Db 824 GTAAATTTCTTTTATGATCTTTGATTTCTGAGAGTCACTGCTGTGTGCAATTAGT 883
 Qy 181 GInGIYAlaSerAspPheLeuLeuLeuGIYLeuSerLeuAspIleAlaProGIYAspMet 200
 Db 884 CAAAGAGTTTCAATTTCTGCTCTCTTGGAAAGCTTTGGACATAGCACAGTGTACATG 943
 Qy 201 LeuAspYsValAlaArgArgLeuSerLeuIleYsHisProGIYCysSerThrMetSer 220
 Db 944 CTTCACAAAGGTGGAGAGAGACTTCTTTAATAAACATCCAGAGTCTCCACATAGAGT 1003
 Qy 221 G1YGIYGIYAsaAlaIleGIuHisLeuAlaYsGIuGIYAsnArgPheHisPheAspIleYs 240
 Db 1004 GTGGGAGAAAGCATTAGAAATTTGGCCAAACAGAAATAGATTTCAATTTGACATCAAA 1063
 Qy 241 ProProLeuHisHisAlaYAsnCysAspPheSerPheThrGIYLeuGIuHisValThr 260
 Db 1064 CTCTCCTTCATCTACTGCTAATAATTTGATTTCTTTTACGAGCTTCAACACCTTACT 1123
 Qy 261 AspYsIleIleMetYsYsGIuYsGIuGIuGIYIleGIuYsGIYGIuIleLeuSer 280
 Db 1124 GATTAATAATTAATGAAGAAAGAAAGAGAGATTTGAAGAGGGCAAAATCTGTCT 1183
 Qy 281 SerAlaAlaAspIleAlaIleThrValGIuHisThrMetAlaCysHisLeuValYsArg 300
 Db 1184 TCAGCAGACAGCATTTGCTGCACAGTACAGACACAAATGCGATTCATCTTGTAAAGA 1243
 Qy 301 ThrHisArgAlaIleLeuPheCysYsGIuArgAspIleuProGIYAsnAsnAlaVal 320
 Db 1244 ACACATGGGGCTATTCGTTTGTGAAGAGAGAGCTTTGATCCCAAAATTAATCAGTA 1303
 Qy 321 LeuValAlaSerGIYGIYAlaAlaSerAsnPheYrIleArgArgAlaLeuGIuIleLeu 340
 Db 1304 CTGTTTGCATCTGTGTGTGTGCGAAGTACTTCTGTATCCCGACAGCTCTGAAATTTTA 1363
 Qy 341 ThrAsnAlaThrGIuYsThrLeuLeuYAspProProArgLeuYsThrAspAsnGIY 360
 Db 1364 ACAAGCAGACAGTGCATCTTGTGTGTCTCTCCACAGATATGACATGATATATGCGC 1423
 Qy 361 IleMetIleAlaThrPheAsnGIYIleGIuArgLeuArgAlaGIYLeuGIYIleLeuHisAsp 380
 Db 1424 ATATATGATGCA-----TGATGTCTCTTGGAGTACATATCAAAAGAGTT 1435
 Qy 381 IleGIuGIYIleArgYrGIuProYsCysProLeuGIYValAspIleSerYsGIuVal 400
 Db 1436 -----TGATGTCTCTTGGAGTACATATCAAAAGAGTT 1471

Qy 401 G1YGIuAlaSerIleYsValProGIuLeuYsMetGIuIle 414
 Db 1472 GGAGAACTTCCATTAAGTACACAGATTTAAATGAGAGTA 1513
 RESULT 9
 BC058172
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Bkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1844)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.P., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McGowan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richard, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A.C., Rodrigues, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22386257
 12477932
 2 (bases 1 to 1844)
 Strausberg, R.
 Direct Submission
 Submitted (15-SEP-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Galtereburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nsl.nih.gov
 Akher, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P.,
 Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legardi, R.,
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Staniford, S., Thomas, P.J., Touchman, J.W.,
 Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 123 Row: d Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 21312463.
 Location/Qualifiers

FEATURES

source

1. 1844
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC:67870 IMAGE:5012054"
/issue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_id="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
222. 1466
/codon_start=1
/product="Unknown (protein for MGC:67870)"
/protein_id="AAH58172.1"
/db_xref="GI:34849664"
/translation="MLMLRTYAGAIIPKPKSKYVGLRFRSVHPRTLSCHKLVIGIET
SCDPTGAIVDENGVIGELALHSGQVHLKTKGIVPPVAQOLHRENIQRIEETLSAS
RIFPSDLAIATIKPGIALSLGVLSLQVNRKPKFPIPIHMEHALITRLTK
VEPFLVILISGKCHLALVQSGDFLLGLSLDIPGDMIDKVARLSILKHPCKT
MSGKAIENLAKDGRFHFTINPPMNKAKNDFSTGLOHTDKLITKEKEBIEKG
QILSSADIAAVOHATACHAKRTHRAILFCQKNLSPANAVALVSGVAASNLVIR
KALEIVANATQCTLCPPRLCTDNGIMIANNGIERLRAGLVLDVIEDIYEPKPL
GIDSRVAAEAIVPRLKMAI"

misc_feature
333. 1397
/note="QRI7, Region: Metal-dependent proteases with possible chaperone activity [posttranslational modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

ORIGIN

Alignment Scores:
Pred. No.: 2.71e-146 Length: 1844
Score: 1835.00 Matches: 352
Percent Similarity: 91.06% Conservative: 25
Best Local Similarity: 85.02% Mismatches: 37
Query Match: 86.35% Indels: 0
DB: 10 Gaps: 0

US-10-649-273-2 (1-414) x BC058172 (1-1844)

QY 1 MetLeuIleLeuThrLysThraAlaGlyValPhePheLysProSerLysArgLysValTyr 20
DB 222 ATGCTAATGTTTAAAGACACGACGACCTATTCACCAAGCCCAAGAAAGTAAATTAAT 281
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
DB 282 GGATTTTAAAGAGTTTAAAGTTCATCCCAAGACTCTCTGTGCATTAACCTGCTCCTG 341
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValAspGluThrGlyValAsnVal 60
DB 342 GGAATTGAACACCACTGTGATGACACAGAGCGCTGTGTGTGATGAATAACGCGGATGTG 401
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
DB 402 CTGGGGAGAGACGTGCACTCCCAACTCAAGGTTCATCTGAATAACGCGGTGATTTCTCT 461
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluValLeuSer 100
DB 462 CCACTAGCTCAACAACCTTCAACAGAAATATTCACAGAAATAGTAGAAGAACTTTTCT 521
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 522 GCCAGTGAATCACCCCAAGCATCTCTCGCAATTCACATCAACATCAACCGGAGACTG 581
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
DB 582 GCCCTAAGCTTAAGAGTCGCTTATCTTAACTTAAGCTTAAGTAATCTTTTAAAG 641
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 642 CCGTTCATCCGATTCATCACTGAGGCTCAAGCATATTAAGCTCAACCAATAAA 701

QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 702 GTAGAATTTCTTTCTTCTAGTTCTTTTGAATTTGGCGGTCATGCTGTGTGGATTAAGTC 761
QY 161 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 762 CAAGGTGTTCCGATTTCTGCTCTTGGGAAGTCTTGTGAATAGCCGAGGCGACAG 821
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlnCysSerThrMetSer 220
DB 822 CTGACCAAGGTGGCAAGAGACTTCTTTAATCAACATCCGAAATGTTCAATGAGT 881
QY 221 GlyGlyLysAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLys 240
DB 882 GGTGAAAAGCTATAGAACATTTGGCCAAAGACGAAATAGATTCATTTACTATCAAT 941
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 942 CCACCTAGCAGAAATGCTAAGATTCGATTTTCTTTTACGGGACTTCAACATATTACT 1001
QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyIleGlnLysGlyGlnIleLeuSer 280
DB 1002 GATTAAGCTTAATTAACAACAAGGAAAAAGAGAGCATTAAGAGGCGCAAAATCCTGTA 1061
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisSerMetAlaCysHisLeuValLysArg 300
DB 1062 TTAGCTGACAGCATTTGCTGCTCGGTACAGATCCAAACGCTGCCACTTGCAGAAAG 1121
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal 320
DB 1122 ACACATCGTGCATTTCTGTTTGGCAAGGAAAAATTTGCTATTCACCTAACCGACGTA 1181
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
DB 1182 TTAGTGTATCTGAGAGTGTTCAGATTAATTGTATCATCCGAAAGCATTTGAAATGTGC 1241
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
DB 1242 GCAAATGCAACGACGACGCTGCTGTGTCCTCCCAAGACTGTGACATGACAAATGGC 1301
QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1302 ATCATGATTTGCATGGAATGGAATTAAGATTAACGTGCTGGCGCTTTTACATGAT 1361
QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400
DB 1362 GTAGAAGCATCCGATATGAAACCAAAATGCTCTTGAATAGACATATCCAGAGAATT 1421
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1422 GCAGAACTGCCATTAAGTACCGGATTAAGTAAAGTGAACCTT 1463

RESULT 10
ARS41929 1416 bp DNA linear PAT 08-OCT-2004
LOCUS ARS41929
DEFINITION Sequence 177 from patent US 6743619.
ACCESSION ARS41929
VERSION ARS41929.1 GI:53934009
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 1416)
Tang, Y. T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F.,
Zhang, J., Zhao, Q. A., Yang, Y., Xue, A. J., Wehrman, T., Wang, J. -R.,
Wang, D. and Dmanac, R. T.
Nucleic acids and polypeptides
Patent: US 6743619-A 177 01-JUN-2004;
location/Qualifiers
1. 1416
/organism="unknown"
/mol_type="genomic DNA"

TITLE
JOURNAL
FEATURES
source

ORIGIN

Alignment Scores:

Pred. No.: 5.9e-139 Length: 1416
Score: 1747.00 Matches: 340
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.71% Mismatches: 0
Query Match: 82.21% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x AR541929 (1-1416)

QY 74 LyeThrclygyllylvalproproalalaglnleuuhisargyluasnlleglnarg 93
DB 280 AGAAGAGGTGGATGTTCTCCAGAGCTCAACAGCTTCAAGAGAAATATTCAACGA 339
QY 94 lilevalglnlualaleuSerlialserglyvalserproserapleuSerlialleala 113
DB 340 ATAGTACMAAGAGCTTTCTGCAGAGGAGTCTTCCAACTGACCTTCAGCAATTGCA 399
QY 114 ThThrllelyspProglyleuAlaleuSerleuglyvalgyleuSerPheSerleuGln 133
DB 400 ACTACCAATAAACAGAGCTTCTTAAAGCTGGAGTGGGCTTATTCATTAGCTTACAG 459
QY 134 LeuvalglyglnleuLylyspProPheleproilehishismetglualahisaleu 153
DB 460 CTGGTAGACAGTTAAABAAAGCCATTCTCCATTATCATATGAGAGCTCATGCACTT 519
QY 154 ThrlleargleuThraenlyvalglnPheProPheleuValleuLeuileserglyly 173
DB 520 ACTATTAGTGTGACCAATTAAGTGAATTTCTTTTATGTTAGTTCTTTGAGAGT 579
QY 174 HiscyleuLeuAlaleuValglnlyvalserapPheleuLeuLeuglylyvserleu 193
DB 580 CACTGCTGTGGATTAGTTCAGAGGAGTTTCAATTTTCTGCTTGGAAAGCTTTTG 639
QY 194 Asp1lalaProgllyaspMetleuAsplyvalalargayleuSerleuileyshts 213
DB 640 GACATGACACAGGTGACATGCTTGAACAGTGGCAAGAACATTTCTTTAATTAACAT 699
QY 214 ProgllyCysSerThMetSerglylylyvalallegluhileuAlalyglnlyasn 233
DB 700 CCAGAGGTCTCCACATGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGAAAT 759
QY 234 ArgPhehishPheapPlelyepProleuHishisalyasaenCyasPheSerPhe 253
DB 760 AGATTTCAATTTTGCATCAACCTCCCTTCATCATGCTAATAAATTTGATTTTCTTT 819
QY 254 ThrglyleuGlnhisvalThraaplylilelleuetylysglylyvglnuglylyle 273
DB 820 ACTGAGCTTCAACGTTACTGATTAATAATTAATTAAGAAAGAAAGAGAAAGTATTT 879
QY 274 GlulysglyglnleuSerSerlialalaspPleuAlaleuPheCyslyvglnargapleu 293
DB 880 GAGAAAGGGGAATTCCTGCTTCAAGCAGCAGCATTCGTGCACAGTACAGCAACATG 939
QY 294 AlaCyehisbleuVallysatgthhisargalaleuPheCyslyvglnargapleu 313
DB 940 GCATGTCATCTTGTGAAGAAACACATCGGGCTATTCTGTTTGTAAAGACAGACTTG 999
QY 314 leuProgllyasapleuAlaleuValalaserglylyvalalaserapPheTyrlle 333
DB 1000 TTAACCTCAAAATATATGATGATGATGATGATGATGATGATGATGATGATGATG 1059
QY 334 ArgArgAlaleuGlnleuThraaplylilelleuetylysglylyvglnuglylyle 353
DB 1060 CGAGAGCTCTGAAATTTTAAACAGCAACAGTGCATTTGTGTGTCTCTCTCC 1119
QY 354 ArgleuCyThraapnglylilemetleuThraapnglylileglnargleuArgala 373
DB 1120 AGACTATGACCTGATATATGATGATGATGATGATGATGATGATGATGATGATGATG 1179
QY 374 GlyleuGlylileuuhisapPleleglnuglylileargTyrglnProlyCyasProleugly 393

DB 1180 GCGTTGGGCAATTTATCATGACATAGAGCATCCGCTATGACCAAAATGCTCTTGGGA 1239

QY 394 ValaapPleleuSerlyglnvalglnlualaserlilelyvalproPleuLeuLyMetGln 413
DB 1240 GTAGACATATCAAAAGAAAGTGGAGAGCTTCCATTAAGTACCACAATTTAAAAATGAG 1299

QY 414 ile 414
DB 1300 ATA 1302

RESULT 11

BC078974 1546 bp mRNA linear ROD 03-AUG-2004
LOCUS BC078974
DEFINITION Rattus norvegicus cDNA clone IMAGE:7111906, partial cde.
ACCESSION BC078974
VERSION BC078974.1 GI:50926879
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

REFERENCE

AUTHORS
1 (bases 1 to 1546)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shemmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,D., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaefer,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McKean,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettelman,M., Madan,A., Young,A.C., Shevchenko,Y.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Boutard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalske,U., Smallus,D.B.,
Scherer,A., Schein,J.B., Jones,S.J. and Marra,M.A.

TITLE

human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL

POBMED
REFERENCES
2 (bases 1 to 1546)
Director MGC Project.
Direct Submission
Submitted (02-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

COMMENT
NIH-MGC Project URL: <http://mgc.mci.nih.gov>
Contract: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLU)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contract: (Dickson, Mark) mcdopax1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNLU at: <http://image.llnl.gov>
Series: Ink Plate: 162 Row: F Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers

|||||
1 ATGAGGCTCAGTACCTACTTATTAGTTGACCATATAAGATATTCCTTTTAGTT 60
168 LeuLeuLeuLeuSerGlyGlyHisCysLeuLeuLeuValGlnGlyValSerAspPheLeu 187
61 CTTTGAATTTCTGGAGGCTGCTGCTGTTGTCATTTAGTTCAAGAGTTTCAATTTTCTG 120
188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
121 CTTCTTGGAAAGCTTTTGGACATGGACCAAGTGCATGCTTGGACAAAGTGGCAAGACA 180
208 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGlnHis 227
181 CTTTCTTAATTAATAACATCCAGAGTGCCTCCACATGATGGTGGGAAAGCCATAGAACAT 240
228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
241 TTGGCCCAACCAAGGAATAGATTTTCATTTTGCATCAACCTCCCTTCATCATGCTAAA 300
248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
301 AATTGTGATTTTCTTTTACTGACCTTCAACACGTTACTGATTAATAATTAATGAAG 360
268 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 287
361 GAAAAAGAGAGAGGTAATTGAGACAGGGCAATCTGCTTCAAGACAGACATGCTGCC 420
288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307
421 ACAGTACAGCACACAAATGGCAATGTCATCTTGTGAAMACAAATCGGGCTAATTCGTTT 480
308 CysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
481 TGTAAAGCAGAGAGCTTTTCTTCAAAATATGCAAGTACTGGTTGCATCTGGTGCTC 540
328 AlSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347
541 GCAAGTACTCTTCTATATCCGACAGGCTCTGGAATTTTACAAAGCAACAGAGGCACT 600
348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 367
601 TTGTGTGTCTCTCTCCACAGACTATGCACTGATTAATGCAATTATGATGCAATGCT 660
368 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGlyTrp 387
661 ATGAAAGACTACCTGCTGCTGCTGCTTTTACATGACATGAAGGATCCGCTATGAA 720
388 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyAlaSerIleLysVal 407
721 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTTGAGAGAGCTTCCATTAAGTA 780
408 ProGlnLeuLysMetGlnIle 414
781 CCACAATTAATAAATGAGAGATA 801

RESULT 13
BX934991 1522 bp mRNA linear VRT 02-FEB-2004
LOCUS BX934991
DEFINITION Gallus gallus finished cDNA, clone CHESTR189114.
ACCESSION BX934991
VERSION BX934991.1 GI:41635519
KEYWORDS
ORGANISM Gallus gallus (chicken)
SOURCE
ORGANISM Gallus gallus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1522)
Boardman, P.B., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.B.,
Croning, M.D.R., Davies, R.H., Francis, M.D., Grafham, D.V.,
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,
Niblett, D., Overton, I.M., Rogers, J., Scott, C.B., Taylor, R.G.,
Tickle, C. and Wilson, S.A.

|||||
37 LysIleValLeuGlyIleGlnThrSerCysAspAspThrAlaAlaValValAspGlu 56
254 AGACTTGTCTGGGATCATCAACCAAGCTGTGACACAGCGCGCGGCTGACAGAG 313
57 ThrGlyAsnValLeuGlyGlnAlaIleHisSerGlnThrGlnValHisLysLysThrGly 76
314 GCGGACACGCTCTGGAGAACCGCTGCAGACCCAAAGAGGTTCCACTCAAGCAGGT 373
77 GlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGln 96
374 GGAATATTTCTTCATGACACAGAGCTTTCACAGGAAGCATCCAGCAAGTAAAG 433
97 GlnAlaLeuSerAlaSerGlyValSerProSerAspLysSerAlaIleAlaThrTrpIle 416
434 GAAGCACTCAGTCCAGTGCAGTCTTCTGTAATGAACCTTGTGCTATTCACACTACAGTG 493
117 LysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGly 136
494 AAACCAAGACTTGCCTGAGCTGAGGAGTGGAGCTGCACATGAGCTTACAGCTGGTGGAC 553
137 GlnLeuLysLysProPheIleProIleHisMetGluAlaHisAlaLeuThrTrpIleArg 156
554 AGGTACCGAAGCCTTTCATACCATTCATCAATGAGAGCTCAGCACTTACATCAGACA 613
157 LeuThrAsnLysValGlnLysPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 176
614 CTGACAGAGCAAGTAAATTTCTTTCTTTAATCTCTCCGAGAGTCACTGCATC 673
177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 196
674 TTGGCAGTACGACGAGGAGTTTCAATTTCTTCTGCTTGGACAGTCCATGATATAGCA 733
197 ProGlyAspMetLeuAspLysValAlaArgArgLysSerLeuIleLysHisProGlnCys 216
734 CCAAGTACATGCTGATTAAGTACGACAAAGAGCTCTTTAGTGAAGACCCGAGTGC 793
217 SerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnLysAsnArgPheHis 236
794 CACGCAATGCTGGGGGAGGCAATAGACACTGCTCAACCGAGACTGGCAACAG 853

|||||
Direct Submission
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: chickens@ems.umist.ac.uk
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
sequencing project.
This sequence is from the
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
from a library constructed by Elizabeth Bosch. cDNA was prepared
from RNA extracted from muscle, normalized, and poly A-tailed.
EcoRI-NciI cut cDNA was then ligated into the vector. Vector:
pBluescript II KS(+); Site 1: EcoRI; Site 2: NciI Host: Escherichia
coli DH10B.
Location/Qualifiers
1..1522
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHESTR189114"
/clone_1fb="CSEORBN11"
/dev_stage="adult"

ALIGNMENT Scores:
Pred. No.: 3.56e-106 Length: 1522
Score: 1362.00 Matches: 256
Percent Similarity: 83.20% Conservative: 54
Best Local Similarity: 68.80% Mismatches: 63
Query Match: 64.09% Indels: 0
DB: 5 Gaps: 0

US-10-649-273-2 (1-414) x BX934991 (1-1522)

Qy	237	theasprileysproplehnihisalalysanCySaARPhSeSePhThGlYeu	256
Db	854	TACATTTCAAGCTTCCATGACACAGTATGTAATCTGTATTTCTTCTCCGACTT	913
Qy	257	GlnHisValThrAspIysIleIleMetIysIysGluIysGluGluGlyIleGluIysGly	276
Db	914	CAGAGCCCTTGCACAAAGCCATTCTTCAGAAAGAAAAGAAAGAGATTCAGAAAGG	973
Qy	277	GlnIleuSerSerSeralalaAspIleAlaIthrValGlnHisIthrMetAlaCysHis	296
Db	974	GAATATCCCTGCTCTGCGTTAAGACATCCCTGCTGTCGACAGACAGTATGCTGCTCAT	1033
Qy	297	LeuValIysaAGThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGln	316
Db	1034	ATTATTCACGCGACACACCCGAGCCATGCTCTTCTGCATGAAAACAGCATATTTATACCA	1093
Qy	317	AsnAsnAlaValIleValAlaSerIysGlyValAlaSerAsnPheTyrlAlaArgAla	336
Db	1094	AAATCGAAGACTCTGCTGTTGTATCAGAGAGAGTTCGAATTAATCAGTATTCAGAAAGGA	1153
Qy	337	LeuGluIleLeuThrAsnAlaIthrGlnCysThrIleuLeuCysProProArgIleuCys	356
Db	1154	CTGCAAGACTCTGCGAAATGCAAAACGGTTTGGCTTTCTGTCTCTCTCTCCACAGGCTGTGC	1213
Qy	357	ThrIrsPasnGlyIleMetIleAlaIthrAsnGlyIleGluIargLeuAlaGlyIleuGly	376
Db	1214	ACCATATATGTGTGTATATATTCAGATGAGATGCAATGTAAGGTGCCGACGAGATGTGCT	1273
Qy	377	IleIeuHisAspIleGluGlyIleArgTyrosGluPProIysCysProLeuGlyValAspIle	396
Db	1274	ATTTTATACAGTACTGATCGATCGATCCGCTACAGAACCAAGCTCCCTTGGATATGATATT	1333
Qy	397	SerIysGluValGlyGluAlaSerIleIleIysValProGlnLeuIys 411	
Db	1334	TCCAAAAGAGTTGAAGAGATTCATCAATGAAGTGCACCAAGACTTATAG 1378	
RESULT 14			
LOCUS	BC038910	1017 bp	mRNA linear ROD 21-OCT-2003
DEFINITION	Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA		
ACCESSION	BC038910		
VERSION	BC038910.1	GI:24433548	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1017)		
AUTHORS	Straubenberg,R.L., Collins,F.S., Grouse,L.H., Dergs,J.G.,		
	Klausner,R.D., Fellins,P.S., Wagner,L., Shennem,C.M., Schuler,G.D.,		
	Altschul,S.F., Zeeberg,B., Butcov,K.H., Schaefer,C.F., Bhat,N.K.,		
	Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,		
	Diatchenko,L., Marinina,K., Farmer,A.A., Riddin,G.W., Hong,L.,		
	Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavent,T.L.,		
	Scapellato,T.E., Brownstein,M.J., Udell,T.B., Toshitsuyki,S.,		
	Carinici,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,		
	Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,		
	McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,		
	Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hilyk,S.W.,		
	Villalton,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,		
	Faney,J., Helton,B., Kettelman,M., Madan,A., Rodriguez,S.,		
	Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,		
	Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,		
	Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,		
	Butterfield,J.S., Krzyzinski,M.I., Skalska,U., Smalins,D.E.,		
	Schneicher,A., Schein,J.B., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length		
	human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 1017)		

AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (25-Oct-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: gcapds-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letlicia Hsiao, Martin Krywninski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeary, Steven Nees, Pawan Pandon, Anna-Lisa Prabhu, Parvaneh Seediq, Jacqueline Schein, Duane Smalusz, Michael Smith, Lorraine Spence, Jeff Stoltz, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.
FEATURES source	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAP Matrix: 86 Row: f Column: 12. Location/Qualifiers 1..1017 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:5053559" /tissue_type="Liver, normal, 5 month old male mouse." /clone_lib="NCI CGAP_L19" /lab_host="DH10B" /note="Vector: pCMV-SPORT6"
ORIGIN	
Alignment Scores:	
Pred. No.:	7 g1e-99 Length: 1017
Score:	1273.00 Matches: 241
Percent Similarity:	92.73% Conservative: 14
Best Local Similarity:	87.64% Mismatch: 20
Query Match:	59.91% Indels: 0
Df:	10 Gaps: 0
US-10-649-273-2 (1-414) x BC038910 (1-1017)	
Oy	140 LybProPhelieProlleHisImecGluJahIsalaleuthrIlleargyleuthrasn 159
Dd	2 AAGCATTCAATCCGATTCAATCATGAGGCTCACGCACGACTTAATAGGTACCAT 61
Oy	160 LybVaGlubheProPhelenuValleuleuileSerGlygylHisCysleuleualleu 179
Dd	62 AAAGTAAATTTCTTTTTTAGTTCTTTGAATTTCTGGCGGTCACTGCTGTGACATTA 121
Oy	180 ValGlnllyValSerarpphelenuleuengilyLysSerleuaarpIIealProglyasp 199
Dd	122 GTCCAAGTGTTTCGAAATTCCTCTCTCTGGGAAGTCTTGGACATAGCAACGGGCAC 181
Oy	200 MetleuaspyValAlaArgArgLeuSerleuileLysHisProglucYseerThmet 219
Dd	182 ATGCTTGAACAAGTGGCAAGAAGCTTTCTTAATCAAACATCCAGAAITGTTCTACATG 241
Oy	220 SerGlyglyLyvalalleGluHisleuAlaLysGlnGlyAsnaArgpheHisPheaspIle 239
Dd	242 AGTGATGAAAAGCTATAAGAACAGTGGCCAAAGACGGAATTAAGATTCATTTACTATC 301
Oy	240 LybProProleuHisHisAlaLysasncYeasrpPheSerRhethtnglyLeuGlnhisVal 259
Dd	302 AATCACCTACCTGACGAATGCTMAAATATGCAATTTTCTTTCACGGACCTCAACATAT 361

Oy		260	TThApbLyIlelilectelysygsluysglugluGlyIleGluyrsylgnileu	279
Oy		261:::	
Db		362	ACTGTATAGCTATAACACACACAGAAAAAGAAAGAGCATTCAGAGGGCGAATCTGC	421
Oy		280	SerSerAlaAlaAspIleAlaAlaThrValGlnHisrthrMeAlaCysHisleuValys	299
Db		422	TCATAGCTGCAGACATTCTCTGCTGGGTACAGCATGCAACAAGCGTGCACCTTGCGAAA	481
Oy		300	AqGTHrHLArGLAlalieleubheCySylsyrglnarngAspleuleuProclnbnbnmla	319
Db		482	AGAACACATCGGCCATATCTGTTTTCGAAGACAGAAAAATTTGCTCTCCAGCTAACGCA	541
Oy		320	VallLeuValAlaSerGlyValAlaAlaSerAnpheTyrlAlargAlaleuGluile	339
Db		542	GTAHTTAGTTGTATCTCGAGAGTGTTCGACGTAACCTTGTACATCCGAAAGCATTTGAAAT	601
Oy		340	LeuthraenAlaThrGlnCysrthrleuEuCysproProPargreucYsrthrapaa	359
Db		602	GTGCAAAATGCACACGACGTGACGTTGTGTGTCACCTCCAAGACTGTGACGTACAT	661
Oy		360	GlylleMelillealatrPhenGlylleGluArgleuArgAlaGlyleuGlylleuHis	379
Db		662	GGCATCATGATTCATGATGAAATGAAATTAAGAATTAACGTGCTGGCGCTTTTACAT	721
Oy		380	ASpIIlegluGlylAlarGyrGluPuPolysCysProleuGlyValAspIIseurysgu	399
Db		722	CATGTAGAGAGACATCCGATATATACCAAAATGTCTCTTGAGATGACATATCCAGAGA	781
Oy		400	ValGlyGluAlaSerIleLysValProGlnleuLysMetGluile	414
Db		782	GTTGCAGAAAGCTGCCATATAAGTACCGCATTTAAATATGCACTT	826
RESULT 15				
LOCUS	BX930963	1558 bp	mRNA	linear VRT 30-MAR-2004
DEFINITION	Gallus gallus finished cDNA, clone CHEST62nl6.			
ACCESSION	BX930963			
VERSION	BX930963.2 GI:46016890			
KEYWORDS				
SOURCE				
ORGANISM	Gallus gallus (chicken)			
REFERENCE				
AUTHORS	Gallus gallus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Ruteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 1558) Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E., Crothing,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V., Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R., Niblett,D., Overton,I.M., Rogers,J., Scott,C.B., Taylor,R.G., Tickle,C. and Wilson,S.A.			
TITLE	Direct Submission			
JOURNAL	Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: chickens@bm.jmst.ac.uk			
COMMENT	On Apr 1, 2004 this sequence version replaced gi:41631491. BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA sequencing project. This sequence is from the BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection, from a library constructed by Elizabeth Bosch. cDNA was prepared from RNA extracted from limbs and poly A-titrated. EcoRI-NotI cut cDNA was then ligated into the vector. Vector: pBluescript II KS(+); Site_1: ScaRI; Site_2: NotI Host: Escherichia coli DHIO.			

FEATURES

Source

```

/organism="Gallus gallus"
/mol_type="rRNA"
/strain="White Leghorn, Hix9"
/db_xref="taxon:9031"
/clone="CHST62n16"
/clone_lib="CSECHL20"
/dev_stage="stage 36"

```

ORIGIN	
Alignment Scores:	
Pred. No.:	4.57e-93
Score:	1208.00
Percent Similarity:	82.54%
Best Local Similarity:	67.46%
Query Match:	56.85%
DB:	5
US-10-649-273-2 (1-414) x BX30963 (1-1556)	
	Length: 1558
	Matches: 228
	Conservative: 51
	Mismatches: 59
	Indels: 0
	Gaps: 0

QY	74	LYETHTRGVYGIYILEVALPROPROLAIAGINGINLEUHSIARGGLASNLILEGINA	93
DB	486	AGAGAGAGGTGAATTAATTCCTCANGTAGCAGCAGGCTTCCAGAGAAAGCATCCAGCAA	545
QY	94	ILEVALINGIUALALEUSERIALASERGIYVALSERPROSERASPLEUSERIALILEALA	113
DB	546	GTAGTAAGGAAGCACTCAGTGCCAGTGAGATTCTGTAAATGAACCTTGCTCTATTGCA	605
QY	114	THRTHTRIELYERPROGIVLEUALALEUSERILEUGLYVALIGLYLEUSERPHESETRLEUGIN	133
DB	606	ACTCAAGTGAACCAAGGACTTGCACTGAGCCTTGAGAGTGGACTGCAGTACAGCTTACAG	665
QY	134	LEUVALIGLYINLEULYERPROPHETILEPROILEHSIEMETGLUALAHISALALEU	153
DB	666	CTGGTGGACAGGTACCAAGACCTTTCATTACCATTATCCATGATGAGGCTTACGCATT	725
QY	154	THRILEARGLLEUTHRASNLVVALIGUNPHEPROPHETLEUVALILEUENILESERGIYGI	173
DB	726	ACCATTCAAGCTGACGAGCCAGTAGTAATTCCTTACGTTCTTTACTCTCCGAGAGT	785
QY	174	HISCYBLEULLEUALALEVALINGIYVALISERASPHETLEULEUGLYLYSSEIU	193
DB	786	CAGTCGATCTTTGGCAGTAGTACGACGAGGAGTTTCAGATTTCCTTGCTTGACAGTCCATA	845
QY	194	ASPLILELAPROGLYASPMETLEUASPLVVALIAAGATGLEUSERLEUENILEYHSIS	213
DB	846	GATTATAGCAACAGGTGATCATCTGATATAGTAGCAAAAGGCTCTTTAGTGAAGAC	905
QY	214	PROGLUCYSESTRTMETSERGIYGIYLYSALALEGINHISLEUALALEGINSIYASN	233
DB	906	CCGAGAGTCCACCGGCAATGCGTGGGGGAGAACCATATAGCAACCTGGCTCAAAACCGAGAC	965
QY	234	ARGPHEHSIPHEASPLILEYERPROPROLEUHSISHISALALYASATCYASPPHSEPHE	253
DB	966	TGGCAACAGTACACTTTCAGACTTCCCATGCAACAGTATCGTAACCTGATTTTCTTTC	1025
QY	254	THRTIYLEUGINHISVALTHRASPLVLEILEMETLYSVYSGIULYUGIUGIUGIYILE	273
DB	1026	TCCGACATTCGAGCCCTTGTCACAAAGCCATTCTTCGAAAGAAAGAAAGAGAGTAT	1085
QY	274	GIULYVSGIUGINILEUSERSERIALALASPLILEALATHTVAILGHNHISHTMET	293
DB	1086	CAAGAAAGGGAAATTCCTGTCTCTGGCTTAAGACATCGCTGCTGCGCACAGCATTAATG	1144
QY	294	ALACYSHISLEUVALYVARGTHHSIARGLAILEUENPHECYLYSGINATGASPLEU	313
DB	1146	GCTGCTGATATTATCCAGCGGACACCGGACCATGCTCTTCGATGATAAAAACAGCAT	1205
QY	314	LEUPROGLINENAMALAVALEUVALIASERGIYVALIALASERASPHETRIELE	333
DB	1206	TTATTATCCAAACATCGCAACTGCTGTGTATACAGAGAGATTTGCAGATATCACATATATC	1265
QY	334	ARGATGALALEUGIUILEUUTHRASNALATHRGINCYSTHTRIEULEUCYERPROPRO	353
DB	1266	AGAAAGAGACTGCGACACTCTGGCAAAAGCAAAACGGTTTGCTTTCTGTCTCTCTCA	1322
QY	354	ASISLEUCYSTHTASPSANGIYILEMECILEALATTPASNGIYILEGLUARGLEUARGLA	373
DB	1326	AGGCTGTGCAACCGATATATGGTATTATGATTCAGATGGCACTTGAAAGGTTGGTGCA	1385
QY	374	GIYLEUGIYILEUHSIASPLILEUGIUGIYILEARGTGIUNPROLYSCYASPROLEUGIY	393

Db 1386 GATGTGTATTATACAGTACTGATGCGCATCGCTATGAACCAAGCTCCCTTGA 1445
Qy 394 ValAspIleSerIysGluValGlyGluAlaSerIleIysValProGlnLeuIys 411
Db 1446 ATTGATATTTCCAAGAGTTGAAGAGGATTCCATCAGATGCCCAAGACTAAGG 1499

Search completed: February 16, 2005, 18:02:36
Job time : 6632 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:09:32 ; Search time 788.4 Seconds

(without alignments)
3108.540 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILTKAGVFFKPSKRVY.....DISKVGSAIKVQLKMEI 414

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Xgapop 10.0, Xgapext 0.5

Delop 6.0, Delext 7.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1h

-O=/cgr2_1/USPTO.spool/US10649273/runat_14022005_114702_16389/app_query.fasta_1.1429

-DB=N_Geneseq_16Dec04 -OPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blousum62 -TRANS=human40.cdi

-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USR=US10649273 @CGN_1_1_1057 @runat_14022005_114702_16389 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGCLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database :

N_Geneseq_16Dec04:*

1: geneeqn1980s:*

2: geneeqn1990s:*

3: geneeqn2000s:*

4: geneeqn2001as:*

5: geneeqn2001bs:*

6: geneeqn2002as:*

7: geneeqn2002bs:*

8: geneeqn2003as:*

9: geneeqn2003bs:*

10: geneeqn2003cs:*

11: geneeqn2003ds:*

12: geneeqn2004as:*

13: geneeqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	2058	6	ABA93268 Human O-s
2	2125	100.0	2197	5	AB276635 DNA encod
3	2125	100.0	2572	8	ABT23207 Human pro
4	2088	98.3	1820	6	AAD46856 Human g1y
5	2088	98.3	1820	8	ACA60887 Human CDN

6	2088	98.3	1821	10	AB557020	Ab557020 cDNA enco
7	1944	91.5	2208	10	ADA52832	Ada52832 Human cod
8	1944	91.5	2890	12	ADQ24627	Adq24627 Human sof
9	1747	82.2	1416	8	ABX70950	Abx70950 Novel hum
10	1549	72.9	1572	6	ABQ75508	Abq75508 Murine si
11	1385	65.2	1526	6	AB76639	Ab76639 DNA encod
12	1001.5	47.1	3358	10	ADB31345	Ad31345 Human d1a
13	714.5	33.6	1601	4	AB124633	Ab124633 Drosophi
14	700.5	32.0	1385	4	AAH15110	Aah15110 Human CDN
15	687.5	32.4	3656	4	AB124632	Ab124632 Drosophi
16	662.5	31.2	1557	3	AAIC38454	Aac38454 Arabidops
17	629	29.6	571	4	AAH12983	Aah12983 Human CDN
18	599	28.2	2734	5	AAAD4622	Aad4622 DNA encod
19	531.5	25.0	4360	6	AAAD48239	Aad48239 Ehrlichia
20	524	24.7	1146	8	ACA26804	Aca26804 Prokaryot
21	502	23.6	1044	8	ACA39102	Aca39102 Prokaryot
22	502	23.6	94750	4	AAAF2851	Aaf2851 Genomic f
23	501	23.6	1053	12	ADL03120	Adl03120 DNA encod
24	500	23.5	1092	6	ABQ90383	Abq90383 M. capsul
25	499.5	23.5	1206	11	ABD02197	Abd02197 Pseudomon
26	494.5	23.3	1032	8	ACA43173	Aca43173 Prokaryot
27	492.5	23.2	1026	4	AA554064	Aas54064 Pseudomon
28	492.5	23.2	1026	8	ACA42146	Aca42146 Prokaryot
29	492.5	23.2	1026	10	ADG73341	Adg73341 P aerugin
30	492.5	23.2	1059	11	ABD02280	Abd02280 Pseudomon
31	490.5	23.1	1026	10	ADG73343	Adg73343 P aerugin
32	484.5	22.8	1014	4	AA552570	Aas52570 E. coli D
33	484.5	22.8	1014	8	ACA51431	Aca51431 Prokaryot
34	484.5	22.8	1014	8	ACA32641	Aca32641 Prokaryot
35	483.5	22.8	1014	8	ACA54107	Aca54107 Prokaryot
36	482.5	22.7	1020	8	ACA44384	Aca44384 Prokaryot
37	482.5	22.7	1074	10	ADP02056	Adp02056 Bacterial
38	481.5	22.7	1014	4	AA556045	Aas56045 Salmonell
39	479	22.5	1315	2	AAQ27645	Aaq27645 Glycophor
40	475.5	22.4	1029	4	AA553309	Aas53309 Haemophil
41	475.5	22.4	1029	8	ACA34150	Aca34150 Prokaryot
42	475.5	22.4	110000	2	AA742063	AA742063 Continuat
43	474.5	22.3	1032	11	ACB96255	Acb96255 Klebsiell
44	472.5	22.2	1014	10	ACF71364	Acf71364 Phototrab
45	472.5	22.2	110000	10	ACF67367_45	Continuation (46 o

ALIGNMENTS

RESULT 1	ABA93268	standard; cDNA; 2058 BP.
ID	ABA93268	
XX	ABA93268;	
AC		
XX		
DT	19-APR-2002	(first entry)
XX		
DE	Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.	
XX		
KM	Human; O-sialoglycoproteinase-like protein; OSGPAP; enzyme; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	/tag= a /product= "O-sialoglycoproteinase-like protein"
FT		110..1354
XX		
PN	CN1318550-A.	
PD		
XX	24-OCT-2001.	
XX		
PF	19-APR-2000; 2000CN-00106834.	
XX		
PR	19-APR-2000; 2000CN-00106834.	
XX		
PA	(SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.	
XX		

PI Mao Y, Xie Y;
 XX WPI: 2002-115090/16.
 DR P-PSDB: ABB05481.
 XX
 PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
 for diagnosing, preventing and treating related diseases.
 XX
 PS Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.
 XX
 CC The present sequence encodes human O-sialoglycoproteinase-like protein
 (OSGRLP). The present invention also describes: (1) the preparation of
 the OSGRLP protein; (2) applying the OSGRLP protein in diagnosis; (3) the
 prevention and/or treatment of related diseases; (4) utilizing the OSGRLP
 protein in screening its agonist, excitomotor and inhibitor and preparing
 an antibody against the OSGRLP protein; and (5) the use of the OSGRLP
 polynucleotide sequences, proteins, agonists, excitomotors, inhibitors
 CC and antibodies in treating diseases related to the abnormal OSGRLP gene
 CC and in preparing the medicine composite for the treatment
 CC
 XX
 SQ Sequence 2058 BP, 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,946-212 Length: 2058
 Score: 2125.00 Matches: 414
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-10-649-273-2 (1-414) x ABA93268 (1-2058)
 QY 1 MetLeuLeuLeuThrLeuThrAlaGlyValaPhePheLeuProSerLeuArgLeuValTyr 20
 DB 110 ATGCTAAATCTTGACCTAAGACGACGAGTCTTTTAAACCATCAAAAGAAAGTTTAT 169
 QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisLeuValLeu 40
 DB 170 GAATTTTAAAGAAAGTTTAAATTTTCTCTGGAACACATTTCTTCATTAATAATGATTG 229
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValaValaAspGluThrGlyAsnVal 60
 DB 230 GGAATTGAAACTAGTGTGATGATGACACAGACTGCTGTGTGTGATGAAGAACTGAAATGTG 289
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuValThrGlyGlyIleValPro 80
 DB 290 TTGGAGAGAGCAATACATTTCCAACTGAAGTTCAATTTAAACAGGTGGATTTGCTCT 349
 QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
 DB 350 CCAGCAGCTCAACAGCTTCAACAGAAATATTTCAACGATATGTCACAGAGCTCTTTCT 409
 QY 101 AlaSerGlyValaSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
 DB 410 GCCAGTGAAGTCTCTCCAAAGTGAACCTTCAGCAATTTGCAATACATAAACACAGACTT 469
 QY 121 AlaLeuSerLeuGlyValaGlyLeuSerPheSerLeuGlnLeuValaGlyIleLeuValys 140
 DB 470 GCTTTAAGCCCTGGAGGTGGCTTTTACATTTAGCTTACAGCTGTGTGGACAGTTAAAG 529
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgPheThrAsnValys 160
 DB 530 CCAATTCATCCCATTCATCATATGAGGCTCATGCACTTACTATTAGGTGACCAATTA 589
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 DB 590 GTAAATTTCTTTTATTTAGTTCTTTTGAATTTCTGAGAGCTCATGCTGTGTGCAATTA 649
 QY 181 GlnGlyValaSerAspPheLeuLeuGlnGlyLeuSerLeuAspIleAlaProGlyAspMet 200
 DB 650 CAAGAGATTTTCAATTTTCTGCTTCTTGAAGAGCTTTTGAACATAGCACAGGTGACATG 709
 QY 201 LeuAspValaAlaArgArgLeuSerLeuIleValHisProGluCysSerThrMetSer 220

DB 710 CTTCACAAAGTGGCAAGAAAGACTTCTTAAATTAACATCCACAGTGCCTCACCATAGT 769
 QY 221 GlyIleValaAlaIleGlnHisLeuAlaValysGlnGlyAsnArgPheHisPheAspIleValys 240
 DB 770 GGTGGGAAAGCCATTAACATTTGGCCAAACAGAAATATGATTTTCAATTTGACATCA 829
 QY 241 ProProLeuHisHisAlaValysAsnCysAspPheSerPheThrGlyLeuGlnHisValaThr 260
 DB 830 CTTCCCTTCATCATCTAATAAATTTGATTTTCTTTTACTGAGACTCAACAGCTTACT 889
 QY 261 AspValIleIleMetLeuValysGlnGlyValysGlnGlyIleGlyValysGlnIleLeuSer 280
 DB 890 GATTAATATATATAGAAAGAAAGAAAGAAAGAAAGTATTCAGAAAGGGCAATCTGCTCT 949
 QY 281 SerAlaAlaAspIleAlaThrValaGlnHisThrMetAlaCysHisLeuValValysArg 300
 DB 950 TCAGCAGCAGACATTCCTCCACAGTACAGACACACAAATGCGATGCTTGTGAAAGA 1009
 QY 301 ThrHisArgAlaIleLeuPheCysValysGlnArgAspLeuLeuProGlnAsnAlaVala 320
 DB 1010 ACAATCGGGCTATTCTGTTTGTATAGCAGAGACTGTGTACTCAAAATTAATGCA 1069
 QY 321 LeuValaIleSerGlyGlyValaAlaSerAsnPhenTyrIleArgArgAlaLeuGluIleLeu 340
 DB 1070 CTGGTTGCATCTGTGTGTGTGCGCAATTAATCTTATATCCGACAGCTCGAAATTTTA 1129
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
 DB 1130 ACAAAAGCAACACAGACCTTGTGTGTCTCCCTCCACAGACTATGACATGATTAAGGC 1189
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
 DB 1190 ATTATATATTCATGAGATGTGATTTGAAGAATCAAGTGTGGCTTTGGCATTTTACATGAC 1249
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValaAspIleSerGlyVala 400
 DB 1250 ATAGAAAGCATCGCATAGAACCAATATGTCCTTGAGATAGCATATCAAAAGAGTT 1309
 QY 401 GlyGluAlaSerIleLeuValProGlnLeuValMetGluIle 414
 DB 1310 GGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1351
 RESULT 2
 ABS76635
 ID ABS76635 standard; DNA; 2197 BP.
 XX
 AC ABS76635;
 XX
 DT 11-DEC-2002 (first entry)
 XX
 DE DNA encoding novel human metalloproteinase MPl.
 XX
 XX Metalloproteinase; MP-1; immune disorder; glutamate transport; cancer;
 motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 liver disease; renal disease; immune disorder; rheumatoid arthritis;
 acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 neurological disorder; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200272751-A2.
 XX
 XX 19-SEP-2002.
 PD
 XX 05-FEB-2002; 2002WO-US003353.
 PF
 XX 05-FEB-2001; 2001US-0266518P.
 PR

KM	gastrointestinal; developmental; vesicle trafficking disorder; infection;
KM	protein-protein interaction; drug-target interaction;
KM	gene expression profile; human; gene; ds.
XX	
OS	Homo sapiens.
XX	
PM	WO2003000844-A2.
PD	
XX	
PF	03-JAN-2003.
XX	
PR	18-JUN-2002; 2002WO-US019360.
XX	
PR	22-JUN-2001; 2001US-0300508P.
XX	
PR	06-JUL-2001; 2001US-0303445P.
PR	13-JUL-2001; 2001US-0305405P.
PR	09-AUG-2001; 2001US-0311442P.
PR	24-AUG-2001; 2001US-0314821P.
PR	29-AUG-2001; 2001US-0315922P.
PR	03-MAY-2002; 2002US-0378205P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;
PI	Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H;
PI	Roselyne JY, Barroso I, Ramkumar J, Griffin JA, Li UX, Yang J;
PI	Thangavealu K, Gietzen KU, Ding L, Baughn MR, Boroweky ML, Yao MG;
PI	Walla NK, Macon PM, Gururajan R, Lee S, Becha SP, Lee SY, Tran UK;
PI	Elliot VS, Luo W, Sprengue MW, Tang YT, Lu Y, Zeharjadian Y;
XX	
DR	WPI. 2003-184039/18.
XX	
DR	P-PSDB; AB026654.
XX	
PT	New isolated human PMOD polypeptide and polynucleotide, useful for
PT	diagnosing, treating and preventing diseases or conditions associated
PT	with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
PT	infections.
XX	
PS	Claim 91; Page 211; 225bp; English.
XX	
CC	The invention relates to an isolated polypeptide comprising: any of 28
CC	sequences of 48-1256 amino acids; a natural amino acid sequence at least
CC	90% identical to the 28 amino acid sequences, 94% identical to a sequence
CC	of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
CC	acids, or 97% identical to a sequence of 242 amino acids, all given in
CC	the specification; or a biologically active or immunogenic fragment of
CC	the isolated polypeptide. The polypeptides and polynucleotides are useful
CC	in diagnosing, treating and preventing diseases or conditions associated
CC	with the decreased expression of protein modification and maintenance
CC	molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
CC	atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
CC	allergies), neurological disorders (e.g. stroke, Parkinson's disease,
CC	epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
CC	endometriosis), developmental, vesicle trafficking disorders, and
CC	infections (e.g. bacterial, viral, parasitic, protozoal). These are also
CC	useful in assessing the effects of exogenous compounds on the expression
CC	of nucleic acid and amino acid sequences of PMOD. The PMOD or its
CC	fragments are useful in screening compounds for effectiveness as agonist
CC	or antagonist of the polypeptides, or in altering the expression of the
CC	target polynucleotide and compounds that specifically bind to or modulate
CC	the activity of the polypeptide. The microarray is useful in monitoring
CC	or measuring protein-protein interactions, drug-target interactions, and
CC	gene expression profiles. This polynucleotide sequence represents the DNA
CC	encoding a human PMOD protein of the invention
XX	
SO	Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;
XX	
Alignment Scores:	
Pred. No.:	2,7e-212
Score:	2125.00
Best Local Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	8
	Gaps: 0
	Matches: 414
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Length: 2572

[illegible]

Dd		1164	ACAAAGCAACACAGTGCACCTTGTTGTGTCCTCCGCCAGACTATGACATGATAATGGC	122
Oy		361	l MecllealATTpasmglyllegluargleualagileuengllylleuH;EASP	380
Dd		1224	ATTATGATTCGATCGAGATGATGATTAAGAAGCTACGTGCTGGCATTTATACATGAC	128
Oy		381	lllegluagllyllearytTgcluppolySvSProlenglyValaspllleSerlySGUval	400
Dd		1284	ATGAGAAGCATCCGCTATGAACCAAATATGCTCTTGAGGTAGACATATCAAGAAGTT	134
Oy		401	GlyGUAlaserlleysValProglneulyMetGlulle	414
Dd		1344	GGAGAAGCTTCATTAATACTACCAATTTAAAATGAGATA	1385
RESULT 4				
AAD46856	ID			
AAD46856	standard; cDNA; 1820 BP.			
XX				
AC	AAD46856;			
XX				
DT	27-JAN-2003 (first entry)			
DE	Human glycoprotease 28472 CDNA.			
XX				
KW	Human; adenosine deaminase; seven transmembrane domain receptor; cancer;			
KW	7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arthritis;			
KW	rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;			
KW	hyperension; ischemic heart disease; obesity; myocardial infarction;			
KW	endothelial cell disorder; Grave's disease; psoriasis; brain disorder;			
KW	Parkinson's disease; Alzheimer's disease; haematopoietic disorder;			
KW	cerebral oedema; metabolic disorder; liver disorder; platelet disorder;			
KW	chromosome mapping; tissue typing; gene therapy; neuroprotective;			
KW	Cystostatic; anorectic; cardiant; haemostatic; gene; ss.			
OS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	146..1390		
FT		/*tag= a		
FT		/product= "Human 28472 protein"		
FT		/note= "This region is specifically claimed as SEQ ID NO:		
FT		6 in claim 1 of the specification"		
FN	WO2002749660-A2.			
XX				
PD	26-SEP-2002.			
XX				
PP	08-NOV-2001; 2001WO-US051427.			
XX				
PR	08-NOV-2000; 2000US-0246768P.			
PR	08-NOV-2000; 2000US-0246772P.			
PR	15-NOV-2000; 2000US-0249185P.			
XX				
PA	(MILL-) MULLENIUM PHARM INC.			
P1	Leiby KR, Kapeller-Libermann R, Glucksmann M;			
DR	WPI; 2002-759898/82.			
XX				
DR	P-PsDB; AAB29234.			
XX				
XX				
PT	New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,			
PT	useful for diagnosing and treating cancer, immune, cardiovascular,			
PT	hematopoietic, brain, pain, metabolic, liver or platelet disorders, and			
PT	in pharmacogenomics.			
XX				
PS	Claim 1; Fig 8; 178bp; English.			
XX				
CC	The present invention relates to novel 38650, 28472, 5495, 65507, 81588			
CC	or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-			
CC	protease or seven transmembrane domain (7TM) receptor family members.			
CC	Sequences of the invention are useful in diagnosing and treating cancer			
CC	or aberrant cellular proliferation and/or differentiation (e.g. colon or			
CC	lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid			

CC	arthritis; diabetes, allergy), heart or cardiovascular disorders (e.g.,
CC	hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
CC	myocardial infarction, thrombus) including endothelial cell disorders
CC	(e.g. psoriasis, Grave's disease), haemopoietic disorders, brain
CC	disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
CC	pain and metabolic disorders (e.g. obesity), liver disorders or platelet
CC	disorders. They are also useful in screening assays, predictive medicine
CC	(e.g. diagnostic assays, prognostic assays, monitoring clinical trials
CC	and pharmacogenetics) and prophylactic and therapeutic methods. The
CC	nucleic acids may also be used in chromosome mapping, tissue typing and
CC	forensic biology and as surrogate markers. Sequences of the invention are
CC	also used in gene therapy. The present sequence is human glycoprotease
CC	28472 CDNA
XX	
SQ	Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
Alignment Scores:	
Pred. No.:	1-24e-208
Score:	2088.00
Percent Similarity:	99.03%
Best Local Similarity:	98.31%
Query Match:	98.26%
DB:	6
Gaps:	0
US-10-649-273-2 (1-414) x AAD46856 (1-1820)	
OY	1 MetLeuileLeuThrLyThrAlaGlyValPhePheLysProSerLyBArgLysValTyr 20
Dd	146 ATGCAATCTTGACCTAAGACTCAGAGAGTTTAAAAACCATTCAAAGAAGTTTAT 205
OY	21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
Dd	206 GAATTTTTAAAGAGTTTAATTTTCATCTCGAACACTATTTCTTCATAAAAATAGATTG 265
OY	41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrArgLysAsnVal 60
Dd	266 GGAATTAAGAACTAGTGTGATATACAGACGCTGCTGTGTGATGAATACTGAAATGTG 325
OY	61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
Dd	326 TTGGGAGAGACATATCATTTCCAAACTGAAGTTTCATTTAAAAACAAGTGGATTGTTC 385
OY	81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Dd	386 CCACGACCTCAACACCTTCACAGAGAAATATTTCAACGAATGTATCAAGAACCTCTTCT 445
OY	101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Dd	446 GCCAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATTTGCACATCAATPAACAAGACCTT 505
OY	121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysVal 140
Dd	506 GCTTTAAACCTGGAGGTGGGCTTATCATTTACCTTACACTGGTAGCACGTTAAAAAG 565
OY	141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Dd	566 CCAATTCATTCATCATCATATGAGAGGCTCAGCACTACTATATAGTTGACCAATAAA 625
OY	161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlyAlaLeuAlaLeuVal 180
Dd	626 GTAGAAATTTCTTTTATGTTCTTTTGAATTTCTGAGAGTACAGTCTGTGTGGCATTAAGT 685
OY	181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Dd	686 CAAGAGTTTCAAGATTTCTGCTTCTTGGAAAGCTTTTGGACATACGACCAAGTGCATG 745
OY	201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer 220
Dd	746 CTTCAGCAACCTCCGCAAGACCTTTCTTAAATAAAACATCCAGAGTCTCCACCATGAAGT 805
OY	221 GlyIleLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Dd	806 GGTTGGAAAGCCATTAAGCAATTTGGCCAACAGAGAAATGATTTCAATTTTGGACATCAA 865

OY		141	ProPheIIePrIleHISHeMeCGLAAlSaIaleThrTlLaXgLeuThrAsnLys	160
Dd		566	CCATTCATTCCCATTCATCATATGAGGGCTCATGCACCTTAATAAGGTGACCAAATGA	625
OY		161	VaIGUpHePrOphLeuValLeuenuIeserGIyHISCySeLeuLeuAlaeuVal	180
Dd		626	GTAGAAATTTCTTTTTTAAAGTTCTTTTGANTTCTGAGGGTCACGTCTGTGGCAATTA	685
OY		181	GlNGIValSerAspPheLeuDeuLeuGIyLysSerLeuAspIlleAlaProGIYaSper	200
Dd		666	CAGAAGATTTCGATATTTCTGCTTTCTTGAAAAGCTTTGGACATAGCACAGTGCATG	745
OY		201	LeuAspLysVaIalaaArgLeuSerLeuIleLysHISproGIUCySerThMetSer	220
Dd		746	CTTGACAAGGTGGCAAGAAAGACTTTCTTTAATAAACATCCAGATGGTCCACCATGAGT	805
OY		221	GIyGIyLVaIalIeGIuHIbleuAlalySGNGIyaNaAgPhenHISpheAspIlleLys	240
Dd		806	GGTGGGAAGCGCATGGAACATTGGCCAAACAGAAATAGATTTCATTTGACATCAA	865
OY		241	ProProlaueHISHISaIalyasasnCySarPheSerPheTrNGIlyLeuGHISvalThr	260
Dd		866	CTCCCTCTGCATCAGCTCTAAAAATTTGATTTTTTTTACTGCACTTCAACAGTTACT	925
OY		261	AspLYSIlelleMeTLySleGIuLySGluGIuGIyIleGIuLVySGIyAlileuSer	280
Dd		926	GATTAABAAATTAATGAABAACAGAAACAAGAGAGATTAAGAAAGGGCCAATCTGTCT	985
OY		281	SeraIlaaIaaspIlleaIalaIathrVaIGlnHISThrMetaIaCySHISleuValLysArg	300
Dd		986	TCAGCAGCAGACATTGCTGCCACAGTACAGCAACAATGGCATGTCATCTGTGAAAAGA	1045
OY		301	ThrHISaRgaIalleuPheCySLySGlnArgAspLeuLeuProGlnAsnaSnAlaVal	320
Dd		1046	AACAATCGGGCTAATCTGTTTTGTGAAGCAGAGACTTTGTAACCTCAAAATTAATGCAGTA	1105
OY		321	LeuValAlaSerSGIyGIyVAlaIaSeranPheTyrlleatArgTaIaleuGlnIleLeu	340
Dd		1106	CTGGTTCATCTGGGGTGGTGCAGAGTAATCTTAATCCGAGAGCTCTGAAAATTTTA	1165
OY		341	ThraenaIaThngInCySThrLeuLeuCysProProProIargLeuCyThraSpasnGIy	360
Dd		1166	ACAAACGACACAGCTGCACCTTGTGTGTCTCTCTCCCAACATHTGACATGATATGGC	1225
OY		361	IleMeCIlealTaTPASnGIyIleGIuArgLeuArgAlagIleuGIyIleleuHISAsp	380
Dd		1226	ATTATGATTCGATGAATGTGATTAAGAAAGTACGCTGCTGGCAATTTTACATGAC	1285
OY		381	IleGIuGIyIleHXGTyrGIuProLycAspProLeuGIyValAspIlleSerLySGIyVal	400
Dd		1286	ATAGAAGCATCCCGCTAATGAACCAAAAGTCTCTTGAGATGACATATCAAAAGAACTT	1345
OY		401	GIyGIuAlaSerIleLysValProGlnLeuLysMetGulle	414
Dd		1346	GGAGAGCTTCATTAAGTAACCAATTAATAAATGAGATA	1387

RESULT 6
 ABS57020 ID ABS57020 standard; cDNA, 1821 BP.

ABS57020;
 30-JAN-2003 (first entry)

cDNA encoding novel human glycoprotease 28472.

KW Cancer; aberrant cell proliferation; aberrant cell differentiation;
 KW breast cancer; ovarian cancer; prostate cancer; colon cancer;
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;
 KW endothelial disorder; hematopoietic disorder; blood vessel disorder;
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;
 KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;

KM		autoimmune disorder; hypertension; atherosclerosis; heart failure;
KM		myocardial infarction; ischemic heart disease; Crohn's disease;
KM		Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
KW		cerebral ischemia; peripheral neuropathy; Alzheimer's disease;
KW		Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.
XX		
OS		Homo sapiens.
FM		
FT	Key	Location/Qualifiers
FT	CDS	147..1391
FT		/tag= a
FT		/product= "Glycoprotease 28472"
FT		/note= "Specifically claimed in claim 1"
XX		
FN		WO200277233-A2.
PD		
XX		
XX		03-OCT-2002.
PF		
XX		08-NOV-2001; 2001WO-US0467724.
PR		
PR		08-NOV-2000; 2000US-0246768P.
PR		08-NOV-2000; 2000US-0246772P.
XX		15-NOV-2000; 2000US-0249185P.
PA		(MILL-) MULLENNIUM PHARM INC.
PI		
XX		Leiby KR, Kapeller-Liebermann R, Glucksmann M;
XX		WPI, 2003-029938/02.
DR		P-PsDB; ABG71162.
XX		
PT		New adenosine deaminase, glycoprotease and seven transmembrane domain
PT		nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
PT		81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
XX		hypertension.
PS		
XX		Claim 2; Fig 8A-B; 178pp; English.
CC		
CC		The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
CC		14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
CC		38650 nucleic acid molecule comprises a sequence encoding adenosine
CC		deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
CC		a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
CC		sequences that encode a human seven transmembrane domain (7TM). The
CC		38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
CC		sequences are useful for diagnosing, preventing or treating a subject
CC		with or at risk of developing a disorder, e.g. cancer or aberrant
CC		cellular proliferation and/or differentiation (e.g. breast, ovarian,
CC		prostate, colon or lung cancer), immune disorders, heart disorders,
CC		cardiovascular disorders, endothelial disorders, hematopoietic disorders,
CC		blood vessel disorders, brain disorders, pain and metabolic disorders,
CC		liver disorders or platelet disorders. These disorders include carcinoma,
CC		sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,
CC		hyperextension, atherosclerosis, heart failure, myocardial infarction,
CC		ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
CC		syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral
CC		neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
CC		cachexia or diabetes. This sequence encodes the novel human glycoprotease
CC		28472
XX		
SQ		Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;
Alignment Scores:		
Pred. No.:	1,24e-208	Length: 1821
Score:	2088.00	Matches: 407
Percent Similarity:	99.03%	Conservative: 3
Best Local Similarity:	98.31%	Mismatches: 4
Query Match:	98.26%	Indels: 0
DB:	13	Gaps: 0
US-10-649-273-2 (1-414)	x	ABSS7020 (1-1821)

```

Db 147 ATGCTATCTTGAACAGACGAGAGGATTTTTTTAAACATCAAAAGGAAGTTTAT 206
Qy 21 GIuPheUaUaSerPheAsnPhenHisProGlyThrLeuPheUeHisIleValIleu 40
Db 207 GAATTTTAAAGATTTTAAATTTTTCATCTGGACACATATTTCTTCAATAAATATGATTTG 266
Qy 41 GIYIleGluThrSerCysAspAspThrAlaIleValIleAspGlyThrGlyAsnVal 60
Db 267 GGAATTTGAAATCTAGTTGATGATGATTCAGACAGCTGCTGGTGGAATGAAATGTTG 326
Qy 61 LeuGIYGIuAlaIleHisSerGlnThrGlyValHisIleUeUySerGlyIYIleValPro 80
Db 327 TTGGAGAGACAAATACATTTCCAAACCTGAAGTTCAATTTAAAAACAGGGGAGTTGCTT 386
Qy 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluIleUeSer 100
Db 387 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATAGTACAGAGAGCTTTCT 446
Qy 101 AlaSerGIYAlaSerProSerAspLeuSerAlaIleAlaThrThrIleUeSerProGlyIleu 120
Db 447 GCCAGTGAAGTCTCTCCAGTGAACCTTCAGCAATTTGCACTACATTAACACAGAGACTT 506
Qy 121 AlaUeSerLeuGIYAlaGIYLeuSerPheSerLeuGlnLeuValGIYGIleUeUySlys 140
Db 507 GCTTTAAGCCTGGAGTGGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAAAAG 566
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaUeThrIleArgLeuThrAsnUys 160
Db 567 CCATTTCATTTCCATTCATCATATGAGAGGCTCATGCACTTACTATTAGGTTAGCCAAATAA 626
Qy 161 ValGIuPheProPheUeValIleUeUeUeSerGIYHisCysLeuUeUeAlaUeVal 180
Db 627 GTAGAAATTTCTTTTATTTAGTTCTTTGATTTCTGAGAGGCTGCTGTGTGCACTTAGTT 686
Qy 181 GlnGIYAlaSerAspPheLeuLeuGIYUySerLeuAspIleAlaProGlyAspMet 200
Db 687 CAAGAGGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGGACATAGCACAGTGTACATG 746
Qy 201 LeuAspUyValAlaArgArgLeuSerIleUyHisProGlyCysSerThrMetSer 220
Db 747 CTTGACAAAGTGGCAGAAAGACTTTCTTTAATAAATCCAGAGTGTCTCCACATGAGT 806
Qy 221 GIYGIYUyAlaIleGIUHisIleUeAlaUyGlnGlyAsnArgPheHisPheAspIleUys 240
Db 807 GGTGGGAAGCCATAGAAACATTTGGCCAAACAGAAATAGATTTCAATTTAACATCAAA 866
Qy 241 ProProUeUHisIleAlaIleUeAsnCysAspPheSerPheThrGlyUeUeGlnHisIleValThr 260
Db 867 CTTCCCTTGCATCATGCTAATAAATTTGATTTTCTTTTACTGACCTTCAACAGCTTACT 926
Qy 261 AspUyIleIleMetUySlysGIUySglUyGlnGlyIleGIUySglYGIleUeUeSer 280
Db 927 GATAAATAATATGAAAACAGGAAACAGAGGAAAGATTTGAGAAAGGGGCAATTCCTGCT 986
Qy 281 SerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisIleUeValUySArg 300
Db 987 TCAGCAGAGCATTGCTGCCACAGTACAGCACAAATGGCAGTGCATCTTGTGAAMAAGA 1046
Qy 301 ThrHisArgAlaIleLeuPheCysUySglUyArgAspLeuLeuProGlyAsnAsnAlaVal 320
Db 1047 ACACATCGGGCTATTTCTGTTTGTAAAGCAGAGAGCTGTATCTTCAATAATATGACAGTA 1106
Qy 321 LeuValAlaSerGIYGIYValAlaSerAsnPhenYrIleArgArgAlaUeUeUeUeUe 340
Db 1107 CTGGTTGCATCTGGTGGTGGTGGCAAGTAATCTTATATCCGAGAGCTGTGGAAATTTTA 1166
Qy 341 ThrAsnAlaThrGlnCysThrLeuUeUeUeUeUeUeUeUeUeUeUeUeUeUeUeUeUe 360
Db 1167 ACAACAGCAACAGTGCATTTGTGTGCTCTCCACAGACTATGCACTATATATGTC 1226
Qy 361 IleMetIleAlaIleTPAsnGIYIleGIUArgLeuArgAlaGIYUeUeUeUeUeUeUeUe 380

```

```

Db 1227 ATTAATGATTCAGTGAATGATTTGAAGAAGTACAGTGGCTTGGCAATTTTACATGAC 1286
Qy 381 IleGIUGIYIleArgThrArgProUySglYValAspIleSerUySglVal 400
Db 1287 ATGAAAGGATCGGCTATGACCAAAATGTCCTTGGAGTTAGACATATCAAAAGAAATT 1346
Qy 401 GIYGIuAlaSerIleUyValProGlnLeuUyMetGluIle 414
Db 1347 GGAAGAGCTTCCATTAATAATACACATTAATAAATGAGATA 1388

RESULT 7
ADA52832
ID ADA52832 standard; cDNA; 2208 BP.
XX
AC ADA52832;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 400.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-0006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
XX
PR 24-JUN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
DR WPI: 2003-395539/38.
DR P-PSDB; ADA54471.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 400; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,15e-193 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 10 Gaps: 1

US-10-649-273-2 (1-414) x ADA52832 (1-2208)
Qy 1 MetLeuIleUeUyThrAlaGIYValPhePheUySglYArgUyValTyr 20
Db 344 ATGCTATCTTGAACAGACGAGAGGATTTTTTTAAACATCAAAAGGAAGTTTAT 403

```

```

QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 404 GAAATTTTAAAGAAATTTTAAATTTTCATCTGGACACATATTTCTTCAATTAATAGATG 463
QY 41 GlyIleGluThrSerCysAspAspThrAlaIleValIleAspGluThrGlyAsnVal 60
Db 464 GGAATTGAACCTAGTTGATGATACAGACAGCTGCTGGTGGTGAAGAACTGGAATG 523
QY 61 LeuGlyIleValIleHisSerGlnThrGluValHisIleLeuYThrGlyIleValPro 80
Db 524 TTGGGAGAAGCAATATACATTTCCCAACTGAGTTCAATTTAAAAACAGTGGATTTCTT 583
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 584 CCAGCACTCAACAGCTTCACAGAGAAATATTCACCAATAGTACAGAAAGCTCTTTCT 643
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleGlyProGlyLeu 120
Db 644 GCCAGTGAAGCTCTCCCAAGTACCTTCACCAATTTGCAATACATTAACCAAGACTT 703
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleLeuYsIys 140
Db 704 GCTTTAAGCTGGAGAGTGGCTTTATCATTTAGCTTACGCTGGTAGACAGTTAAAGAG 763
QY 141 ProPheIleProIleHisIleMetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnYs 160
Db 764 CCATTCATTCCTCATTCATCATATGAGGCTCATGACCTTACATTAAGTTAGCCATATAA 823
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuVal 180
Db 824 GTAGAAATTTCTTTTATGATCTTTGATTTCTGAGAGCTCACTGCTGTGGCATTAAGT 883
QY 181 GlnGlyValSerAspPheLeuLeuGlyIleYsSerLeuAspIleAlaProGlyYsMet 200
Db 884 CAAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTTTTGACATAGCACCAAGTACATG 943
QY 201 LeuAspYsValAlaArgArgLeuSerLeuIleYsHisProGlyCysSerThrMetSer 220
Db 944 CTTCACAAAGGTGGCAAGACCTTCCTTATTAACATCCAGAGTCTCCACCATGAGT 1003
QY 221 GlyGlyIleValIleGlnHisIleGlnHisIleGlnHisIleGlnHisIleYs 240
Db 1004 GGTGGGAAACCCTATAGAACTTTGGCCAAACAGGAATTAAGATTTCAATTTGACATCAA 1063
QY 241 ProProLeuHisIleHisIleYsAsnCysAspPheSerPheThrGlyLeuGlnHisIleValThr 260
Db 1064 CTTCCCTTGATCATGCTGCTAAATAATGCAATTTCTTTTACCTGACCTTCAACACCTTACT 1123
QY 261 AspYsIleIleMetCysIleGlyGluYsGluGlyIleGlyIleGlyGlyIleLeuSer 280
Db 1124 GATTAATTAATTAATGAAGAAAGAAAGAAAGGATTAAGAAAGGAGGCAAAATCTGTCT 1183
QY 281 SerAlaIleAspIleAlaIleAlaThrValGlnHisIleThrMetAlaCysHisIleValYsArg 300
Db 1184 TCAGCAGCAGACATGCTGCGACAGTACAGACACAAATGCGATGCTTGTGAAAAG 1243
QY 301 ThrHisArgAlaIleLeuPheCysYsGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 1244 ACACATCGGGCTATTCGTTTGTGTTAGCAGAGACATGTTACCTCAAAATTAATGCACTA 1303
QY 321 LeuValAlaSerGlyValAlaIleSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340
Db 1304 CTGGTTCATCTGGTGGTGGTGGCAAGTAATCTTGTAATCCGACAGCTCTGAAAATTTTA 1363
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlyLeuCysThrAspAsnGly 360
Db 1364 ACAACCCACACACAGTCACTTTGTGTGCTCTCTCCACACATATGACATGATATGAC 1423
QY 361 IleMetIleAlaThrPasnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1424 ATATATGATGCA----- 1435
QY 381 IleGluGlyIleArgGlyGluProYsCysProLeuGlyValAspIleSerYsGluVal 400

```

```

Db 1436 -----TCATGTCTCTGGAGTACACATATCAAAAGATT 1471
QY 401 GlyGluAlaSerIleYsValProGlnLeuYsMetGluIle 414
Db 1472 GGAGAAAGCTTCATTAAGTACCAATTAATAATGAGATA 1513

RESULT 8
ADQ24627
ID ADQ24627 standard; DNA; 2890 BP.
XX
AC ADQ24627;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnick A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 7447; 210bp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;

Alignment Scores:
Pred. No.: Length: 2890
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: Gaps: 1

US-10-649-273-2 (1-414) x ADQ24627 (1-2890)
QY 1 MetLeuIleLeuThrIleYsThrAlaGlyValPhePheYsProSerYsArgYsValYr 20
Db 1001 ATGCTAATCTGACTAAGACTGACAGAGTTTATTTTAAACATCAAAAGAAAGATTAT 1060

```



```

QY      21  GluPheLeuArgSerPheLeuPheHisProGlyThrLeuPheLeuHisLysIleValIleu 40
Db      1061 GAATTTTAAAGAACTTTTAAATTTTCATCTCGAAACAATATTTCTTCATTAATAATAGTATTTG
QY      41  GlyIleGluThrSerCysAspAspThrAlaAlaValIleAspGluThrGlyValenVal 60
Db      1121 GGAAATTAAGTAAGTGTGAAGTAACAGACAGCTGTGTGGTGAAGAACTGGAATATGTG 1180
QY      61  LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
Db      1181 TTGGGAGAGCAATACATTCCTCCAACTGAAGTTCATTTAAACAAGTGGGATTTTCTTCT 1240
QY      81  ProAlaIaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db      1241 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAAATAGTACAGAGAGCTTTTCT 1300
QY      101  AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db      1301 GCCAGTGAAGTCTCTCCAAAGTACCTTCACCAATGTGCACTACCAATAAACAGAGACTT 1360
QY      121  AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuLysLys 140
Db      1361 GCTTTAAAGCTGGAGAGTGGCTTATTCATTAGCTTACAGCTGGTGAAGACAGTTAAAG 1420
QY      141  ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db      1421 CCATTCAATCCCATTCATCATATGAGAGCTCATGACCTTACATTAAGTTGACCAATAA 1480
QY      161  ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db      1481 GTAAATAATTCCTTTTATGATCTTTTGATTTCTGAGAGCTCACTGTCTGTGGCATTAAGT 1540
QY      181  GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyLysMet 200
Db      1541 CAAGGAGTTTCAGATTTTCTGTCTTGAAAGTCTTTTGACATAGCACAGGTGACATG 1600
QY      201  LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer 220
Db      1601 CTTCACAAAGGTGGCAAGAAAGCTTCCTTAATTAACATCCAGAGTCTCCACATAGAGT 1660
QY      221  GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db      1661 GGTGGGAAAGCCATAGAAATTTGGCCAAACAGAAATAGATTTTCATTTGACATCAAA 1720
QY      241  ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisAlaThr 260
Db      1721 CTTCCCTGTGATTCATGCTTAAATAATTTGATTTTCTTTTACTGCACTTCAACACCTTACT 1780
QY      261  AspLysIleIleMetLysLysGluLysGluGlyIleGluLysGlyGlnIleLeuSer 280
Db      1781 GATTAATAATTAATGAAAAGAAAAGAAAAGAAAGGATTTGAGAGGGGCAAAATCTGTCT 1840
QY      281  SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
Db      1841 TCACACACAGACATTTGCTGCCACAGTACAGACACAAATGGATGTCATCTGTGAAAAGA 1900
QY      301  ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db      1901 ACACATGGGGCTATTCGTTTGTGTAAGCAGAGACTTGTAACTCCAAATAATATCCAGTA 1960
QY      321  LeuValAlaSerGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340
Db      1961 CTGCTTGCAATGTGGTGGTGGTGCAGAGTACCTGTAACCGAGAGCTCTGGAATTTTA 2020
QY      341  ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuGlyCysThrAspAsnGly 360
Db      2021 ACAAAACCAACACAGTGCACCTTTGTGTGCTCTCTCCACAGATATGACATATATGGC 2080
QY      361  IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db      2081 AATTATGATTCGA----- 2092
QY      381  IleGluGlyIleArgTyrGluPheCysProLeuGlyValAlaAspIleSerIleGluVal 400

```

```

Db      2093 -----TCATGCTCTCTGTGAGTGAAGATATCAAAAGAGTT 2128
QY      401  GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db      2129 GGAGAGCTTCATTAAGTACCAATTAATAATGAGAGATA 2170

RESULT 9
ABX70950
ID      ABX70950 standard; cDNA; 1416 BP.
XX
AC      ABX70950;
XX
XX
XX      05-MAR-2003 (first entry)
DE      Novel human cDNA sequence #175.
XX
XX      Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW      Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW      neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW      autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW      insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
KW      ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW      fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW      coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW      Crohn's disease; anaphylaxis; proliferation; chemoplastic;
KW      differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW      haemostatic; antiinflammatory; expressed sequence tag; EST.
XX
XX      Homo sapiens.
OS
XX
XX      WO200281731-A2.
PN
XX
XX      17-OCT-2002.
PD
XX
XX      29-JAN-2002; 2002WO-US001222.
PF
XX
XX      30-JAN-2001; 2001US-00774528.
PR
XX
XX      (HSE-) HSEQ INC.
PA      (GOOD/) GOODRICH R W.
XX
XX      Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QN, Ren F;
PI      Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
XX
XX      WPI; 2003-058563/05.
DR
XX
XX      Claim 1; Page; 612pp; English.
XX
XX      This invention relates to the cDNA sequences encoding an isolated novel
XX      human polypeptide. The protein encoded by the nucleic acid of the
XX      invention is useful for treating central and peripheral nervous system
XX      diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
XX      lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
XX      Alzheimer's disease); autoimmune disease (e.g. systemic lupus
XX      erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
XX      ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
XX      ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
XX      osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
XX      trauma); lung or liver fibrosis; reperfusion injury in various tissues;
XX      bacterial, viral or fungal infections; allergic conditions such as
XX      allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
XX      cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
XX      disease, anaphylaxis). The protein may be used to inhibit the growth,
XX      infection or function of infectious agents such as bacteria, fungi,
XX      viruses, or to effect bodily characteristics, biorhythms or circadian
XX      cycles of rhythms. The protein may also have
XX      proliferation/differentiation, stem cell growth factor, haematopoiesis,
XX      regulation, immune stimulating or suppressing, chemotactic/chemokinetic,

```

CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
CC activities. The cDNA sequences of the invention are useful for expressing
CC recombinant protein for analysis. The present sequence represents a novel
CC human cDNA sequence of the invention, this sequence is an expressed
CC sequence tag (EST) and was identified using subtractive hybridisation

XX Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,31e-173	Length:	1416
Score:	1747.00	Matches:	340
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.71%	Mismatches:	0
Query Match:	82.21%	Indels:	0
	8	Gaps:	0

US-10-649-273-2 (1-414) x ABX70950 (1-1416)

```

QY 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
Db 280 AGAACAGGTGGATTGTTCTCTCCAGCAGCTCAACAGCTTCCAGAGAAATTTCAAGA 339
QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 340 ATACTACAGAAAGCTCTTCTGCCAGTGCAGTCTCCAGAGTCACTCTCAGCAATTGCA 399
QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
Db 400 ACTACCAATAAACCGAGCTGCTTAAAGCCGGAGGAGGCTTATTCATTAGCTTACG 459
QY 134 LeuValGlyGlnLeuLysLeuPhePheIleProIleHisIleMetGluAlaHisAlaLeu 153
Db 460 CTGGTAGGACAGTTAAABAAAGCCATTCCATTCAATGAGAGCTCATGACTT 519
QY 154 ThrIleArgLeuThrAsnLysValGlnPheProPheLeuValLeuLeuIleSerGly 173
Db 520 ACTATTAGTTGACCAATAAAGTAAATTTCTTTTATTTGTTGTTGTTGAGGT 579
QY 174 HisGlyLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193
Db 580 CACTGCTGTTGGATTAGTTCAGAGAGTTCAGATTTCTGCTTCTTGAAAGCTTTG 639
QY 194 AspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHis 213
Db 640 GACATTAGCAGGAGTGCATGCTGACAAAGTGGCAAGAGCTTTCTTAATAAACA 699
QY 214 ProGluCysSerThrMetSerGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsn 233
Db 700 CCAGAGTCTCCACCATAGTGGTGGAAAGCCATAGAACTTTGGCCAAACAGAAAT 759
QY 234 ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPhe 253
Db 760 AGATTTCATTGTGACATCAAACTCCCTGTCATCATGCTAAAAATGATTTTCTTTT 819
QY 254 ThrGlyLeuGlnHisValThrAspLysIleIleMetLysGluLysGlnGlnGlyLe 273
Db 820 ACTGAGCTTCAACAGTTACTGATTAATAATTAAGAAAGGAAAAAGAGAAAGTAT 879
QY 274 GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMet 293
Db 880 GAGAAAGGGGAAATCTCTCTTCCAGCAGACATGCTGCCACAGTACAGCACAATG 939
QY 294 AlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 313
Db 940 GCATGTCATCTTGAAAAGAAACATCGGCTATTTGTTTGAACAGAGACACTTG 999
QY 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaLysAsnPhePheIle 333
Db 1000 TTACTCTCAATAATATGAGTCTGTTGCATCTGTTGTCGCAAGTAACCTTCATATTC 1059
QY 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
Db 1060 CGCAGAGCTCTGAAAATTTTAACAAACGCAACAGTCACTTGTGTGTCTCTCTCC 1119

```

```

QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAla 373
Db 1120 AGACTTGCATGCACTAATAGCATTAATGATTGCAATGGAATGTAATGAAACATCGTCT 1179
QY 374 GlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysProLeuGly 393
Db 1180 GGCTTGGCATTTTATCATGACATGAGAGCATCGGCTATGAAACCAAAATGCTCTTGA 1239
QY 394 ValAspIleSerIleGlnValGlyGluAlaSerIleLysValProGlnLeuLysMetGlu 413
Db 1240 GTAGACATATCAAAAAGAGTTGAGAAAGCTTCCATTAAGTACACAAATTTAAATGAG 1299
QY 414 Ile 414
Db 1300 ATG 1302

RESULT 10
AB075508
AB075508 standard; DNA; 1572 BP.
AB075508;
07-NOV-2002 (first entry)
Murine sialoglycoprotease-like gene sequence SEQ ID NO.7.
Murine sialoglycoprotease-like gene sequence SEQ ID NO.7.
Murine, mouse; protease; calcium activated neutral protease type 5;
CAPNS; trypsinase 4; sialoglycoprotease; enzyme; genetic disease;
neurological; neuropsychological; psychotic illness; transgenic animal;
gene; db.
Mus musculus.
WO200245491-A2.
13-JUN-2002.
05-DEC-2001; 2001WO-US046405.
06-DEC-2000; 2000US-0251803P.
06-DEC-2000; 2000US-0251820P.
13-DEC-2000; 2000US-0255971P.
(DELT-) DELTAGEN INC.
Allen KD, , Leviten MW;
WPI; 2002-657389/70.
Novel transgenic animal, comprising a disruption in protease target gene,
is useful for identifying agents that ameliorates a phenotype associated
with a disruption in a protease target gene.
Example 3; Fig 7; 62pp; English.
The present invention describes a non-human transgenic animal (I)
comprising a disruption in a protease target gene (PG) selected from
calcium activated neutral protease type 5 (CAPNS) gene, trypsinase 4 gene
and sialoglycoprotease-like gene. Also described is a targeting construct
(II), comprising a first polynucleotide sequence homologous to at least a
first portion of PG, a second polynucleotide sequence homologous to at
least a second portion of PG and a selectable marker. (II) is useful for
producing a transgenic mouse comprising a disruption in a protease target
gene, by introducing (II) into a cell, introducing the cell into a
blastocyst, implanting the resulting blastocyst into a pseudopregnant
mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
and breeding the chimeric mouse to produce the transgenic mouse. (I) is
useful for identifying an agent that modulates the expression or function
of a protease target gene, by administering an agent to (I) and
determining whether the expression or function of the disrupted protease
target gene in (I) is modulated. (I) is also useful for testing the
efficacy of proposed genetic and pharmacological therapies for human

```


CC genetic diseases, such as neurological, neuropsychological or psychotic
 CC illness. The present sequence represents murine sialolipoprotease-like
 CC gene sequence, which is used in an example from the present invention

XX Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

Alignment Scores:

Pred. No.:	3.76e-152	Length:	1572
Score:	1549.00	Matches:	326
Percent Similarity:	84.50%	Conservative:	23
Best Local Similarity:	78.93%	Mismatches:	60
Query Match:	72.89%	Indels:	7
	6	Gaps:	2

US-10-649-273-2 (1-414) x ABQ75508 (1-1572)

```

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
DB 5 ATGCTAATGTTAAAGAGAGACAGACGACCTATTCCCAAGCCCCCAAGGTAAGTTTAT 64
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
DB 65 GGAATTTTAAAGAAAGATTATTAAGTTCATCCAGAACTCTCTCTTGTCAATAAATGGATCTG 124
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValAlaAspGlyThrGlyValAsnVal 60
DB 125 GGATTTGAAACCACTGTGATGACACAGAGCGCTGTGTGATGAAATCGGAATGTG 184
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyIleValPro 80
DB 185 CTGGGGAGAGACAGCTGCACTCCCAAACTCAGGTTTCATGAAAACGGTGGGATTTCTTCT 244
QY 81 ProAlaAlaGlnGlnLeuHisArgGlyAsnIleGlnAlaGlyIleValGlnAlaLeuSer 100
DB 245 CCACTTACTCAACAACCTTCAAGAAAAATTTCAACAAATAGTAAGAAAGAACTCTTCT 304
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 305 GCCGTGGAATCACCCCAAGCGATTTCTCAGCAATTCGCACTACCAACCGGAGACTG 364
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleLeuLysLys 140
DB 365 GCCCTAAGTTTGGAGTTGGCTTATCCTTGGCTTACAGCTAGTAATCAGTTTAAARS 424
QY 141 ProPheIleProIleHisMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 425 CCATTTCAT-CCGATTTCATCAGATGAGGCTCACGACGACATATTAAGCTCACCAATAA 483
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 484 GTAGATTN-CTTTTATTAAGTCTTATTAATCTGCGGCTCAGCTGCTGTGAMRTTAGTC 542
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 543 CAAAGTGTTCCTCGATTCCTCTCTCTGGGAAGCTTTTGGACATAGCACAGGCAACTG 602
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlnCysSerThrMetSer 220
DB 603 CTTCACAAAGGTGGCAGAAAGACTTCTTAAATCAAAACATCCAGAAATGTTTCAATGAGT 662
QY 221 GlyValLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 663 GGTGGAAGGCTAAGAAACAGTGTGGCAAAAGCGAAATAGATTTCATTTTATCTATCAT 722
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 723 CCACTTATGCAAGATGCTTAAGATTTGCGATTTTTCACAGGCACTTCAACATATTACT 782
QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyGlyIleGlyLysGlyIleLeuSer 280
DB 783 GATTAAGCTAATTAACACACAAAGAAAGAAAGAAAGGCAATTGAGAAAGGCAAAATCTGTCA 842
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
  
```

```

DB 843 TAGGTGAGACATTGCTCTGCGGTACAGATGCAACAGCGTGCACCTTGCGAAAGA 902
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnAlaGAspLeuLeuProGlnHisAsnAlaVal 320
DB 903 ACACATCGGCTAATCTGTGTTGCAAGAAAAATTTGCTCTTCACACTAACGACGTA 962
QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340
DB 963 TTAGTTGATCTGAGAGGTGTTCAGATTAACCTTGATCATCGAAACATTTGAAATGTTC 1022
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
DB 1023 GCAAAATGCAACGACAGCAAGGTTGTGTGTCCACT-TCAAGACTGTGCACTGCAATGGG 1081
QY 361 IleMetIleAlaIleTrpAsnGlyIleGluValGluAlaGlyIleLeuHisAsn 380
DB 1082 CATATGATTCATGGAATGGAATTTGGAATTAACGCTGTGCTTGCCCTTTTACATGA 1141
QY 380 P-IleGluGlyIleArgGlyGlyProLysCysProLeuGlyVal--AspIleSerLysG 399
DB 1142 TGATTAAGACATTCGCTTATTAACCAAAATGCTCTTGATGAGGACATTAATCCGAAAG 1201
QY 399 LuValGlyGluAlaSerIleLysValProGlnLeu 410
DB 1202 AGTTTGGCAGA---AGCTTGCCCATTAATAAAGTTA 1233
RESULT 11
ABST76639 standard; DNA; 1526 BP.
AC ABST76639;
XX 11-DEC-2002 (first entry)
DE DNA encoding novel human metalloprotease MPI fragment #1.
XX
KW Metalloprotease; MP-1; immune disorder; glucamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200272751-A2.
XX
PD 19-SEP-2002.
XX
PF 05-FEB-2002; 2002MO-US003353.
XX
PR 05-FEB-2001; 2001US-0266518P.
XX
PR 10-APR-2001; 2001US-0282814P.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Chen J, Feder J, Nelson TC, Duclos F, Kyvstek S,
XX
DR MPI, 2002-723329/78.
XX
DR P-PSDB; ABG96487.
XX
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
XX
PT treating, or ameliorating diseases associated with aberrant
XX
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
XX
PT neurological disorders.
XX
PS Disclosure; Page 462-463; 473pp; English.
  
```

CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This sequence represents a
 CC metalloprotease MPI polynucleotide

XX Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5.9e-135	Length:	1526
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	65.18%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x ABS76639 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 167
 DB 1 ATGAGGCTCATGACCTACTTACATAGCTTGACCAATAAGATGATTTCTTTTACTT 60
 QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
 DB 61 CTTTTCATTTCTGAGGCTGACCTGCTGTTGGCATTAAGTTCAGAGTTTCAGATTTTCG 120
 QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
 DB 121 CTTCTTGGAAAGCTTTGGACATGACACAGGTGACATGCTTGACAGGTGGCAAGAAAGA 180
 QY 208 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 227
 DB 181 CTTTCTTTAATAAATCCAGAGTCTCCACATGATGTTGGAAAGCCATAGAACAT 240
 QY 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
 DB 241 TTGGCCAAACAGAGAAATAGATTTTCATTTTGACATCAAACTCCCTTCATCATCTATAA 300
 QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
 DB 301 AATTGTGATTTTCTTTACTGAGACTTCACACCGTTACTGATTAATAATATGAAAAAG 360
 QY 268 GluLysGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 287
 DB 361 GAAAAAAGAGAGGTATTGAGAAAGGGCAAAATCCGCTTTCAGCGACATGCTGCC 420
 QY 288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307
 DB 421 ACAGTACAGACACAAATGCGATCTCTCTGAAAAGAACATCGGGCTATTTCTGTTT 480
 QY 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 327
 DB 481 TGTAAAGAGAGAGACTGTTAAGCTCAAAATATGACAGTACGTGTCATCGTGTCTC 540
 QY 328 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347
 DB 541 GCAAGTAACTTATATATCCGAGAGCTCTGGAATTTTAAACAAACGCAACACAGTGCCT 600
 QY 348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleTrpAsnGly 367
 DB 601 TTGTTGTCTCTCTCCCAAGCTATGACATGATTAATGCGATTATGATTCAGAAATGAT 660

QY 368 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgArgGlu 387
 DB 661 ATTGAAGACCTACGCTCTGCTGGCATTTTACATGACATTAAGACATCCGCTATGAA 720
 QY 388 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 407
 DB 721 CCAAAATGCTCTTGGAGTACATATCAAAAAGAGTTGGAGAGCTTCATTAAGTA 780
 QY 408 ProGlnLeuLysMetGluIle 414
 DB 781 CCACAAATTAATAAATGAGATA 801
 RESULT 12
 ADE31345/C
 ID ADE31345 standard; DNA; 3358 BP.
 XX
 AC ADE31345;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human diagnostic and therapeutic polynucleotide (dthp), SEQ ID No 100.
 XX
 KW diagnostic and therapeutic polynucleotide; dthp; antiarteriosclerotic;
 KW antiinflammatory; cerebroprotective; antilipemic; antidiabetic;
 KW immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;
 KW osteopathic; antiarthritic; antirheumatic; cytostatic; hepatotropic;
 KW virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;
 KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
 KW thrombolytic; anticoagulant; anorectic; vasotropic; antitumor;
 KW gene therapy; protein replacement therapy; human; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003062376-A2.
 XX
 PD 31-JUL-2003.
 XX
 PF 13-JAN-2003; 2003WO-US001096.
 XX
 PR 16-JAN-2002; 2002US-0349384P.
 XX
 PR 17-JAN-2002; 2002US-0349413P.
 XX
 PR 17-JAN-2002; 2002US-0349946P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GB, Jackson JT,
 PI Yu JY, Tusson O, Yap PE, Ameshey SR, Dam TC, Liu TP, Gerstein EH;
 PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RV, Uraahka ME;
 PI Kriecnam SR, Kolluru V, Panesar IS;
 XX
 DR WPI; 2003-636732/60.
 DR P-PSDB; ADE31156.
 XX
 PT New human diagnostic and therapeutic polynucleotides and polypeptides,
 PT useful for diagnosing, treating or preventing e.g. leukemia, brain
 PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke
 PT or Alzheimer's.
 XX
 PS Claim 1; SEQ ID NO 100; 634pp; English.
 XX
 CC The invention relates to a novel isolated human diagnostic and
 CC therapeutic polynucleotide (designated dthp). The novel dthp
 CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
 CC base pairs fully defined in the specification; a polynucleotide
 CC comprising a naturally occurring polynucleotide sequence at least 90%
 CC identical to the dthp polynucleotide; a polynucleotide complementary to
 CC the dthp polynucleotide or its polynucleotide which is at least 90%
 CC identical; or an RNA equivalent of any of the polynucleotides mentioned
 CC above. The dthp polynucleotides have the following activities:
 CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipemic,
 CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
 CC tranquilizer, osteopathic, antiarthritic, antirheumatic, cytostatic,

Dh	1512	GCATTTTCATGACATATGAAAGCATCCGCTATGAACCAAAATGTCCTTTGGAGTAGACA	1453
Oy	396	leSeLyvGluValGlyGluAlaSerIleYvAlProGlnLeuYswetGluIle	414
Dh	1452	TATCAAAAGAGTTGGAGAGCTTCATTAAGAATGCCAATTAATAATGAGATA	1397
RESULT 13			
ABL24633			
ID	ABL24633	standard; DNA; 1601 BP.	
AC	ABL24633;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 25372.		
XX			
KM	Drosophila; developmental biology; cell signalling; insecticide;		
XX	pharmaceutical; gene; ds.		
OS			
XX	Drosophila melanogaster.		
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001..		
XX			
PF	23-MAR-2001; 2001WO-US009231.		
XX			
PR	23-MAR-2000; 2000US-0191637P.		
XX	11-JUL-2000; 2000US-00614150.		
PA	(PEKE) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI; 2001-656860/75.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signaling and cell-cell		
XX	interactions.		
XX			
XX	Claim 1; SEQ ID NO 25372; 21pp + Sequence Listing; English.		
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (AB146176-AB130511), expressed DNA		
CC	sequences (AB101840-AB146175) and the encoded proteins (ABB57737-		
CC	ABB572072). The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX			
SQ	Sequence 1601 BP; 344 A; 477 C; 441 G; 339 T; 0 U; 0 Other;		
Alignment Scores:			
	Pred. No.:	1.51e-64	Length: 1601
	Score:	714.50	Matches: 156
	Percent Similarity:	58.22%	Conservative: 60
	Best Local Similarity:	42.05%	Mismatches: 140
	Query Match:	33.62%	Indels: 15
		Gaps: 6	
US-10-649-273-2 (1-414) x ABL24633 (1-1601)			
Oy	39	ValleuGly11eGluThrSerCyAspAspThrAlaAlaValValAspGluThrGly	58
Dh	79	GTCTGGGCGATCGAGACTCTCTGGAGACACACGGGCGATCGCATGCGACACACGGGC	138
Oy	59	AsnValleuGlyGluAlaIleHisCserGlnThrGluValHisLeuYsThrGlyGlyIle	78
Dh	139	CGAATGTATGTGCATATGTCGAGAGTCGACACAGAGATTCACACCCGCTATGAGGCGATT	198

QY	79	ValProProAlaAlaGlnGlnLeuHisArgGluAmaIleGlnArgIleValGlnGluA	98
DB	199	ATTCGCCCGCCAGGCCCCAGGACCTTACACCGCGCCGACGACTGACCTCCAGCGCTGC	258
QY	99	LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro	118
DB	259	ATGAGAGCGCGCCAAATTGAAGCCGGAACCACTGACGGCCATGCGGGTGACACGGGTCC	318
QY	119	GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerPheSerLeuGlnLeuValGlyLeu	138
DB	319	GGACTCGCGCTGAGTTTCTGGTGGCGCGTGGCGCTTGGACGGCACCTGGCCCGCTCG	378
QY	139	LysLysProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThr	158
DB	379	CAGAAGCCCCCTCCCGCTCCGCTTCAACATGAGAGGGCCACGCGCTCAGGCCCCGATGGAA	438
QY	159	Asn-----LysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlyLeu	176
DB	439	CATCCGAGACAGATGGCTACCCCTCTCTGCTTGGCCAGCGGGCCACCTGTACG	498
QY	177	LeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAla	196
DB	499	TTGGTGGTGGCTAAGCGCCCGCGTCCGCTTACGCTTCTTGGCCAAACACTGATATCG	558
QY	197	ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys	216
DB	559	CCCGCGAGAGCTTTGACAAAGATCGGTGGCGCATACCGCTTGCACATTCGCGGAGTAC	618
QY	217	SerThrMetSerGlyGlyLysAlaIleGlnHisIleLeuAlaLys--GlnGlyAsnArgPhe	235
DB	619	CGCTTGTGAACGGAGGAGCGGGCCATTAGACATGCGCCGACGCTGCGACGACATCGCTG	678
QY	236	HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	255
DB	679	GGCTACAGACTTCCCGCTGCGCTCCAGCAGCGGGAATGCAATTCAGCTTCCGCGCG	738
QY	256	LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlnGlyIleGluLys	275
DB	739	ATCAAGAACAATCTCTCCGCGGCATCCGAGCGCGGACGCGGCGGACGACACCTCCG	798
QY	276	GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetLacys	295
DB	799	GACGGACATCAGACAACTACGCGGCACTTTCGCGCGGCTGCGCTGTGTACGCGG	858
QY	296	HisLeuValLysArgThrHisArgAlaIleLeuPheCys-----LysGlnArgAspLeu	313
DB	859	CACCTGATGACACCGCACGCGACGCGCAATTAGTACTCTCTCCGCCGACAGGACGCTC	918
QY	314	LeuProGlnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheThrIle	333
DB	919	TTCCGGTACACCCCGCCACACTGTGTATGTCGCGCGGTGTGGCAACAACAGATCCATA	978
QY	334	ArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro	353
DB	979	TACGCCAACATCAGCAACATCTGCGCGCGGAGTATGGCTGCAGAGCTTCCGTCAACGAAG	1038
QY	354	ArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuAlaGla	373
DB	1039	CGGTACTCTCGACACACGCGCTATGATCCCTGGCACGCGGCTTGAACACTG-----	1092
QY	374	GlyLeuGlyIleLeuHisAspIleGlu---GlyIleArgTyrGluProLysCysProLeu	392
DB	1093	-----CTCGACGAGTAAAGAGCCGACGCGCTACGACTAGAC-----	1151
QY	393	GlyValAspIleSerLysGluValGlyGluA	403
DB	1132	AGCATGTATATCCAGGCGACGCGGATTCGCC	1164

AAH15110/c
 ID AAH15110 standard; cDNA; 1385 BP.
 AC AAH15110;

QY 204 ----- 204
Db 207 TTTCATTATTCCTTGCATCTTTCGTTTCACAGATTAAATTTATGACTTAA 148
QY 205 ----- 205
Db 147 AATATGTTTCTTGATAGTGGCAGAGACCTTCTTAATAAATCAATCCAGAGTGTCC 88
QY 218 ThrmetserglylylvalailegyluhsleualalygnglyAsnhrpheiisPhe 237
Db 87 ACCATGATGTGTGGAAAGCCATAGACATTTGGCCAAACAGAAATGATTTTCAATTT 28
QY 238 AspilleuYpProPoleuHishisAla 246
Db 27 GACATCAACCTCCCTTCATCATGCT 1
RESULT 15
ABL24632/c
ID ABL24632 standard; DNA; 3656 BP.
AC ABL24632;
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 25369.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li FWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 25369; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABBS72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3656 BP; 1035 A; 829 C; 874 G; 918 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 3.5e-61 Length: 3656
Score: 687.50 Matches: 157
Percent Similarity: 55.64% Conservative: 60
Best Local Similarity: 40.26% Mismatches: 139
Query Match: 32.35% Indels: 34
DB: 4 Gaps: 7

US-10-649-273-2 (1-414) x ABL24632 (1-3656)
QY 39 ValLeuGIYlIeGluThrSerCySaAspThrAlaAlaValAlaAspGluThrGly 58
Db 2578 GTCTTGGGATGAGACCTCTCTCCAGACGAGCGGATCGCTGTGGACACACGGGC 2519
QY 59 AsnValLeuGIYlValAlaIleHisSerGlnThrGluValHis----- 72
Db 2518 CCAATGATTCGCAATGTGCTGGATCGCAACAGAGTTCACACCGCGTGAAGTCTACA 2459
QY 73 -----LeuYThr-GLYglYlIleValPr 80
Db 2458 TCTCCGATCGCGTCCCAATGATGATCAATGATGATGATGATGATGATGATGATGAT 2399
QY 80 OFPAlaIaGlnGlnLeuHieArgGluBsnIleGlnArgGlnIleValGlnGlnAlaLeu 100
Db 2398 GCCCAGGCGCCAGACCTTCAACCGCCGCGCATCGAGTCCGCTTACAGCGCTGATGGA 2339
QY 100 rAlaSerGIYalIserProSerAspLeuSerAlaIleAlaThrThrIleYsProGIYle 120
Db 2338 GCGGCGGCAATTGAGCGCGGACCACTGACGCGCATCGCGTGAACAGCGCTCCGCACT 2279
QY 120 uAlaLeuSerLeuGIYalGIYleuSerPheSerLeuGlnLeuValGIYlIleuYsly 140
Db 2278 GCCGCTGAGTTGCTGTGGCGGTGGCGCTTGCACGCGCACTTGGCCCGCTGCAGAA 2219
QY 140 sProPheIleProIleHieImerGluAlaHisAlaLeuThrIleArgLeuThrAsn-- 159
Db 2218 GCCCTGTGCGCTTGCACACATGGAGCGGCGGCTGCAGCGCCGATGGAATCC 2159
QY 160 ----LySValGIYpPhePheLeuValLeuLeuIleSerGIYlYHieCySLeuLeuAl 178
Db 2158 GGAGCAGATCGGCTACCTTCTCTGTGCTTGGCAGCGGCGGCACTGTAGTTGGT 2099
QY 178 aLeuValGlnGIYalIserAspPheLeuLeuLeuGIYlySerLeuAspIleAlaProGI 198
Db 2098 GGTGGCTAACGGCCCGGTGGCTTACGCTTCTTGGCAAACTGTGATGATGCGCCGG 2039
QY 198 YAspMetLeuAspIYValAlaArgArgLeuSerLeuIleYsHieProGIYsSerTh 218
Db 2038 CGAGGCTTGTGACAGATCGGCGGCTGACGCTGACATTTCTCCGAGTACCGCTT 1979
QY 218 rHetserglylylvalailegyluhsleualalyS-----GlnGIYAsnhrpheiisP 237
Db 1978 GTGAAACGAGGACCGGCGCATTTGACATCCGCGCCAGCTGGCCAGCGATCCGCTGCTA 1919
QY 237 eAspIleuYpProPoleuHishisAlaIleAlaIleAlaIleAlaIleAlaIleAla 257
Db 1918 CGAGTTTCCGCTGCCATCTTCCAGAGGAGGAACTCACTTCACTTCCGCGGATCA 1859
QY 257 nHisValThrAspIYleIleMetYleuGIYlYsGIYlYleGIYlYleGIYlYleGIY 277
Db 1858 GAACAACCTCTTCCGGGCGCATCCGAGCGGCGGAGCGGAGCGGAGCACTCCGAGCG 1799
QY 277 nIleLeuSerSerAlaAlaAspIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 297
Db 1798 AATCATTCAGCAACTACGCGGACTTCTGCGCGGCGCTGCTGCTGCTGCTGCTGCTG 1739
QY 297 uValIYsArgThrHieArgAlaIleuPheCyS-----LyGIYAspLeuPur 315
Db 1738 GATGACCGGACGAGCGGCGGCAATTCATCTCTCCGCGGCGGACGAGCACTCTTCGG 1679
QY 315 oGlnAsnAsnAlaValLeuValAlaSerGIYlYalAlaIleAspAspPheThrIleArgAr 335
Db 1678 TGACACCCCGCCACACCTGTATGTCGGCGGTGGCCAAACAGATGCCATATACCG 1619
QY 335 gAlaLeuGluIleuThrAsnAlaThrGlnCySThrLeuLeuCySProProProArgIle 355
Db 1618 CACATCCCACTTCTGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1559
QY 355 uCyThrAspAsnGIYleMetIleAlaThrAsnGIYlIleGIYlYsArgLeuAlaIle 375
Db 1558 CTGCTCCACACGCGGCTATATGCTCTGCGCGGCTGAGCACTG----- 1510

Qy 375 ucglyleleuHisaSPIlegiu---GlyilearGTyrGluProlysCysProleuGlyVa 394
Db 1509 -----CTGCAGGATTAAGAGGCCAGCAGCGGCTACGACTACGAC-----AGCAT 1466
Qy 394 IAspIleSerIysGluValGlyGluAla 403
Db 1465 TGATATCCAGGGCAGCGCGGATTCGCC 1438

Search completed: February 16, 2005, 13:58:31
Job time : 834.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:19:37 ; Search time 246.15 seconds
(without alignments)
2752.056 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 2125
Sequence: 1 MLILITKAGVFPKPSKRVY.....DISKEVGEASIKVQLKMEI 414

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-Q=/cg2n_1/USPTO.spool/US10649273/runat_14022005_114704_16425/app.query.fasta_1.1429
-DB=Issued_Patents_NA -OFMT=faaetap -SUFFIX=rni -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MITLEN=0 -MAMLEN=200000000
-USER=US10649273 @CGEN 1.1.292 @runat_14022005_114704_16425 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued_Patents_NA:*
- 1: /cg2n_6/ptodata/1/ina/5A.COMB.seq:*
 - 2: /cg2n_6/ptodata/1/ina/5B.COMB.seq:*
 - 3: /cg2n_6/ptodata/1/ina/6A.COMB.seq:*
 - 4: /cg2n_6/ptodata/1/ina/6B.COMB.seq:*
 - 5: /cg2n_6/ptodata/1/ina/PTUS.COMB.seq:*
 - 6: /cg2n_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	2197	4 US-10-067-443-1	Sequence 1, Appli
2	2090.5	98.4	1387	4 US-10-067-443-21	Sequence 21, Appl
3	1747.5	82.2	1416	4 US-09-774-528-177	Sequence 177, App
4	1385	65.2	1526	4 US-10-067-443-23	Sequence 23, Appl
5	1186.5	55.8	14364	4 US-10-067-443-20	Sequence 20, Appl
6	502	23.6	94750	4 US-09-596-002-38	Sequence 38, Appl
7	501	23.6	1053	4 US-09-540-236-806	Sequence 806, App
8	499.5	23.5	1206	4 US-09-252-991A-801	Sequence 801, App
9	492.5	23.2	1059	4 US-09-253-991A-084	Sequence 884, App
10	482.5	22.7	1074	4 US-09-543-681A-2341	Sequence 2341, Ap
11	482	22.7	996	4 US-09-902-540-6612	Sequence 6612, Ap
12	482	22.7	2582	4 US-09-902-540-503	Sequence 503, App

13	479	22.5	1315	1 US-08-087-797-1	Sequence 1, Appli
C 14	475.5	22.4	1830121	4 US-09-557-884-1	Sequence 1, Appli
C 15	475.5	22.4	1830121	4 US-09-643-990A-1	Sequence 1, Appli
C 16	474.5	22.3	1032	4 US-09-489-039A-2050	Sequence 2050, Ap
17	472	22.2	1092	4 US-09-107-532A-2955	Sequence 2955, Ap
18	461	21.7	1008	4 US-08-987-121A-5	Sequence 5, Appli
19	461	21.7	1011	3 US-08-987-121A-5	Sequence 3, Appli
20	457	21.5	1006	3 US-08-961-083-51	Sequence 51, Appli
21	457	21.5	1006	4 US-09-536-784-51	Sequence 51, Appli
C 22	457	21.5	10974	3 US-08-961-527-214	Sequence 214, App
23	455	21.4	1011	3 US-09-066-512-1	Sequence 1, Appli
24	454	21.4	1101	4 US-09-134-000C-1551	Sequence 1551, Ap
25	453	21.3	1011	4 US-09-583-110-2196	Sequence 2196, Ap
26	448	21.1	1011	4 US-09-107-433-1618	Sequence 1618, App
27	443	20.8	1107	3 US-09-134-001C-1072	Sequence 1072, Ap
28	433	20.4	1026	3 US-09-149-624-1	Sequence 1, Appli
C 29	424	20.0	15249	4 US-08-956-171E-102	Sequence 102, App
C 30	424	20.0	15249	4 US-08-781-986A-102	Sequence 102, App
31	406.5	19.1	640681	4 US-09-790-988-1	Sequence 1, Appli
32	406.5	19.1	1230025	4 US-09-198-452A-1	Sequence 1, Appli
33	406.5	19.1	1230230	4 US-09-438-185A-1	Sequence 1, Appli
34	404.5	19.0	1155	4 US-09-602-777A-147	Sequence 147, App
C 35	404	19.0	3064	3 US-09-221-017B-794	Sequence 794, App
C 36	379.5	17.9	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 37	379.5	17.9	4411529	3 US-09-103-840A-1	Sequence 1, Appli
38	369.5	17.4	645	4 US-09-328-352-483	Sequence 483, App
39	362	17.0	876	4 US-09-724-623-19	Sequence 19, Appli
40	357	16.8	580073	4 US-08-545-528D-1	Sequence 1, Appli
C 41	333.5	15.7	42325	4 US-08-311-731A-131	Sequence 131, App
C 42	327.5	15.4	3993	4 US-09-710-279-3985	Sequence 3985, App
C 43	319	15.0	822	4 US-09-710-279-727	Sequence 727, App
C 44	317	14.9	36941	4 US-08-311-731A-130	Sequence 130, App
45	313	14.7	1664976	4 US-08-916-421B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-067-443-1

Alignment Scores:
Pred. No.: 5.72e-261
Score: 2125.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4
US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

Length: 2197
Matches: 414
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

QY 1 MetLeu1leuThrIlySerThraGlyValPhePheIysProSerIyAsrGlyValIlyr 20
 Db 231 ATGCTAATCTTGAATAAGACTGACAGAGTTTATTTTAAACATCAAAAAGAAAGTTTAT 290
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIlyValIleu 40
 Db 291 GAATTTTAAAGAACTTTTAAATTTTCACTCTCGAAACACTATTTCTTCATTAATAATAGTATTTG 350
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValIleAspGluThrGlyIysAsnAl 60
 Db 351 GGAATTAATAACTAGTATGATGATACAGACCTGCTGGTGAAGAACTGGAATATGTG 410
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleuIySerThrGlyIyIleValPro 80
 Db 411 TTGGGAAGAAGCAATACATTTCCCAATGAGATTCATTTAAAAACAGGTGGAGTTTCTT 470
 QY 81 ProAlaIleGlnGluLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
 Db 471 CCAGCACTCAACAGCTTCACAGAGAAATATTCACCAATATGATCAAGAAAGCTCTTCT 530
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIyProGlyIleu 120
 Db 531 GCCAGTGAAGCTCTCCAGAGTACCTTCAGCAATGCACTACCATTAACACAGACTT 590
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuIySlys 140
 Db 591 GCTTTAAGCTCCGGAGTGGCTTATCATTTAGCTTACGCTGGTGAAGACAGTTAAAAAG 650
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIyS 160
 Db 651 CCATTCATTCCTCATTCATATGAGAGCTCATGCACTTACATTAAGTTAGCCAAATAA 710
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 Db 711 GTAGAAATTTCTTTTATGATCTTTGATTTCTGAGGTCACTGCTGTGGCATTAAGTT 770
 QY 181 GlnGlyValSerAspPheLeuLeuGlyIySerSerLeuAspIleAlaProGlyAspMet 200
 Db 771 CAAGGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGGACATGACCCAGGTACATG 830
 QY 201 LeuAspIyValAlaArgArgLeuSerLeuIleIyHisProGlyCysSerThrMetSer 220
 Db 831 CTTGACAAAGGTGGCAAGAGACTTTCTTAAATAAACAATCCAGAGTGTCCACCATGTAGT 890
 QY 221 GlyGlyIyValAlaIleGluHisLeuAlaIyGlnGlyAsnIyAsnIyPheAspIleIyS 240
 Db 891 GGTGGGAAGGCCAATAGCAATTTTGCCCAACAGGAATAATGATTTTCATTTGACATCAAA 950
 QY 241 ProProLeuHisHisAlaIyAsnGlyAspPheSerPheThrGlyLeuGlnHisIyValThr 260
 Db 951 CTTCCCTTGCAATGCTAATAAATGTGATTTTCTTTTACCTGAGCTTCACACGTTACT 1010
 QY 261 AspIyValIleIleMetIySlysGlyIyLeuGlnGlyIleGlyIySlysGlyIleuSer 280
 Db 1011 GATTAATAATTAATGAAGAAAAGAAAAGAGAGATTAAGAGAGGCAATCTGTCT 1070
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisIleThrMetAlaCysHisIleuValIyArg 300
 Db 1071 TCAGCACAAGCATGCTGCGCAGTACAGCACCAAGGCAATGTCATCTGTGAAAGA 1130
 QY 301 ThrHisArgAlaIleLeuPheCysIyGlnArgArgLeuProGlnAsnAsnAlaAlaI 320
 Db 1131 ACAATGAGGCTAATCTGTTTGTGTAAGCAGAGAGACTGTTTACCTCAAAATTAATCACTA 1190
 QY 321 LeuValAlaSerGlyIyValAlaSerAsnPheIyIleArgArgAlaLeuGluIleu 340
 Db 1191 CTGGTTGCACTGTGTGTGTGCGCAAGTACTTCTATATCCGACAGAGCTCTGGAATTTT 1250
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlyLeuCysThrAspAsnIy 360
 Db 1251 ACAAAACCAACAGAGTCACTTTGTGTGTCTCTCCACAGCTATGACATGATATGAGC 1310
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380

Db 1311 ATTAGATTCGATGAAATGATGATGAAGAAGCTACAGCTGCGCTGGCATTTTACATGAC 1370
 QY 381 IleGluGlyIleArgIyGlnIyProIySlysCysProLeuGlyValAspIleSerIySlysVal 400
 Db 1371 ATAGAAAGGATCGCTATGAAACCAAAATGTCTTGGAGTGAACATATCAAAAAGAGTT 1430
 QY 401 GlyIleAlaSerIleIyValProGlnLeuIySmetGluIle 414
 Db 1431 GGAGAACTTCCATTAAGTACCAATTAATAATGAGATA 1472
 RESULT 2
 US-10-067-443-21
 ; Sequence 21, Application US/10067443
 ; Patent No. 6642041
 ; GENERAL INFORMATION:
 ; APPLICANT: Bristol-Myers Squibb Company
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
 ; TITLE OF INVENTION: SPINAL CORD, MF-1
 ; FILE REFERENCE: D00073 NP
 ; CURRENT APPLICATION NUMBER: US/10/067,443
 ; CURRENT FILING DATE: 2002-02-05
 ; PRIOR APPLICATION NUMBER: US 60/266,518
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: US 60/282,814
 ; PRIOR FILING DATE: 2001-04-10
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 21
 ; LENGTH: 1387
 ; TYPE: DNA
 ; ORGANISM: homo sapiens
 US-10-067-443-21
 Alignment Scores:
 Pred. No.: 6,62e-257 Length: 1387
 Score: 2090.50 Matches: 412
 Percent Similarity: 93.85% Conservative: 0
 Best Local Similarity: 93.85% Mismatches: 2
 Query Match: 98.38% Indels: 25
 Ds: 4 Gaps: 1
 US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)
 QY 1 MetLeu1leuThrIlySerThraGlyValPhePheIysProSerIyAsrGlyValIlyr 20
 Db 24 ATGCTAATCTTGAATAAGACTGACAGAGTTTATTTTAAACATCAAAAAGAAAGTTTAT 83
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIlyValIleu 40
 Db 84 GAATTTTAAAGAACTTTTAAATTTTCACTCTCGAAACACTATTTCTTCATTAATAATAGTATTTG 143
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValIleAspGluThrGlyIySlysAsnAl 60
 Db 144 GGAATTTAAAGAACTTTTAAATTTTCACTCTCGAAACACTATTTCTTCATTAATAATAGTATTTG 203
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleuIySerThrGlyIyIleValPro 80
 Db 204 TTGGGAAGAAGCAATACATTTCCCAATGAGATTCATTTAAAAACAGGTGGAGATTTGCT 263
 QY 81 ProAlaIleGlnGluLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
 Db 264 CCAGAGCTCAACAGCTTCACAGAGAAATATTCACCAAGATGTGTAAGAGCTTTTCT 323
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIySlysProGlyIleu 120
 Db 324 GCCAGTGAAGTCTCTCCAGAGTACCTTCAGAAATGCACTACCATTAACCAAGACTT 383
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuIySlys 140
 Db 384 GCTTTAAGCTCGAGAGTGGCTTATCATTTAGCTTACAGCTGGTGAAGACAGTTAAAAAG 443
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIyS 160

Dd	444	CCATTCAATCCCATTCATCATATGAGAGGCTATGACCTTACTATTAGTTAGGTCACATATAA	503
Oy	161	ValGluPheProPheLeuValLeuLeuIleSerGlyGlnHisCysLeuLeuValLeuVal	180
Dd	504	GTAGAAATTCCTTTTATTTAGTCTTTTGATTTCTTGAGAGGTCACTGTCTGTGGCATTAGTT	563
Oy	181	GlnGlyValIleSerPheLeuLeuLeuGlyLeuSerIleuAspIleAlaProGlyAspMet	200
Dd	564	CAGGAGATTCAGATTTTCTGTTCTTGGAAGCTTTTGAAATAGACACACAGGTACATAG	623
Oy	201	LeuAspIysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer	220
Dd	624	CTTGACAAAGTGGCAAGAAAGCTTTCTTTAATAAACAATCCAGAGTGTCTCCACCATAGT	683
Oy	221	GlyGlyLysValAlaIleGlnHisLeuValAlaLysGlnGlyAsnArgPheHisPheAspIleLys	240
Dd	684	GGTGGGAAGGCATGAGCATTTGGCCCAACAGAGAAATAGATTCAATTTTGACATCAAA	743
Oy	241	ProProLeuHisHisAlaLysAsnGlyAsnProPheSerPheThrGlyLeuGlnHisValThr	260
Dd	744	CTCTCCCTTGACATCAAGCTTAATAATGTGATTTTCTTTTAACTGACCTTCAACACGTACT	803
Oy	261	AspLysIleIleMetLysLysGlyLysGlnGlyIle-----	273
Dd	804	GATATAATATATATGAAAAAGAAAGAAAGAGATATATTTCTAATTAATGAAGTTGAA	863
Oy	274	-----GluLys	275
Dd	864	CAGATAAATATATCTGATGTGCTAAATAATAGCTGCTCATTTCTGACAGTATAGAAAG	923
Oy	276	GlyGlnIleLeuSerSerIleAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys	295
Dd	924	GGGCAAAATCTGTCTTCAGACAGACAAATGCTGTGCAACAGTAACAGACAAATGGCATGT	983
Oy	296	HisIleValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro	315
Dd	984	CATCTTGTAAGAAAGACACATCGGCTATTTCTGTTTGTGAACAGAGAACTTGTTACTT	1043
Oy	316	GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAsnAsnPheTyrIleArgArg	335
Dd	1044	CAAAATATATGACGATCTGGTTCGATCTGGTGTGTGTGCAAGTAATCTTATATCCGACGA	1103
Oy	336	AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	355
Dd	1104	GCTCTGGAATATTTTAAACAACGCCAACACAGTGCATCTGTGTGTGTCTCTCCACAGACTA	1163
Oy	356	CysThrArgAsnGlyIleMetIleAlaThrAsnGlyIleGlyAsnArgLeuArgAlaGlyLeu	375
Dd	1164	TGCATGTATTAATGGCATTAATGATTCACAGAAATGATTTGAAGAATACGTACGTGTGGCTTG	1223
Oy	376	GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp	395
Dd	1224	GGCATTTTACATGATGATGAAGAGCATCCGCTATGAACCAAAATGTCCCTTGGAGTAAAC	1283
Oy	396	IleSerLysGlyValAlaGlyLysAlaSerIleLysValProGlnLeuLysMetGluIle	414
Dd	1284	ATATCAAAAGAAAGTGAAGAGCTTCATTAATAAGTACACACATTAATAATGGAATGA	1340

[illegible]

Best Local Similarity: 32.20% Mismatches: 2
Query Match: 55.84% Indels: 657
DB: 4 Gaps: 4

US-10-649-273-2 (1-414) x US-10-067-443-20 (1-14364)

QY 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
DB 10623 AGAACAGGTGGAGATTGTTCCCGACAGCTCAACAGCTTCACAGAAATTTACAGA 10682
QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
DB 10683 ATAGTACAGAGGCTCTTTCGACAGGAGCTCTCCAAAGTACCTCCAGCAATTCAC 10742
QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
DB 10743 ACTACCAATTAACAGACAGCTGCTTTAAGCTGGAGAGGCTTATATTAAGCTTACG 10802
QY 134 LeuValGlyGlnLeuLysLysPheProPheIleProIleHisIleMetGluAlaHisAlaLeu 153
DB 10803 CTGTAGAGACAGTTAAATAAACCATTCATTCCTCATATGAGGCTCATGCACTT 10862
QY 154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173
DB 10863 ACTATTAGTTGACCAATTAAGTAGAATTTCTTTTATGTTCTTTGATTTCTGAGGT 10922
QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193
DB 10923 CACTGTCGTGGCATTTAGTTCAGAGAGTTCAGATTTCTGCTTCATGAAAGTCTTGG 10982
QY 194 AspIleAlaProGlyAspMetLeuAspLysVal----- 204
DB 10983 GACATGACCAAGGTGACATGCTTGACAGGT -AATTAGAATTATTTCTCCATTTCTT 11041
QY 204 ----- 204
DB 11042 TTTGTTATGTTGTCATTTCACTAAGTAGCAATGATGCTACACCATTCACCTAAA 11101
QY 204 ----- 204
DB 11102 TATTTCTGAATTTTATCTAGTAACTGAAAAAATTCAATATGCTGAGAAAAATAGA 11161
QY 204 ----- 204
DB 11162 AAGAGTAGTACACATTTTATATTTCTAGCCTTTCTTAATAAATGTAAGAGGTCAT 11221
QY 204 ----- 204
DB 11222 ATCTGTACATAAAGCTGAATAGTTTGCAATACATTAATATTTTGCATAATATGT 11281
QY 204 ----- 204
DB 11282 ATGTGAAGAAGCGTCTTCGTAACATACTACGAAAAAAGTTAAATAAGAAATAT 11341
QY 204 ----- 204
DB 11342 ATATAGATTAAACATTAAGACATTAAGATGCAATGCAAGATTAAATCAACAATTACT 11401
QY 204 ----- 204
DB 11402 TACACACAGACAGAGGTCCTCCCGCCCTTTGTTTATGAATATCTACAGAGGCTACT 11461
QY 204 ----- 204
DB 11462 GCCATATATAGAAAAACAACAACAACAACAACAACAACAACAACAACAACAACA 11521
QY 204 ----- 204
DB 11522 TAGGAATATAGACAGATTCTTATTAATTAAGCTTACATTAAGCAGTTATGTCACT 11581
QY 204 ----- 204
DB 11582 TCAGCCCATTTTCCAAACAATAGAGAGCAACAATAGACAGGGGAGTAGATTGGCTCT 11641

QY 204 ----- 204
DB 11642 TATTTGGGTCATCATTAAGAACAGGTTGTCTGCTTAAGTAATATGCTATATGTC 11701
QY 204 ----- 204
DB 11702 TATATTGGCAAGATATAGCATGTTTATTCATTCAGGGGTTTTTTGTTTGTAGTAAT 11761
QY 204 ----- 204
DB 11762 TTTCAATTATTTCTTTGCACTTTTCTTCACAGTATTAATTATTAAGCTTAAAA 11821
QY 205 ----- 11AARGAGLeuSerLeuIleLysHisProGlyCysSer 217
DB 11822 AATATGTTCTTTGATAGGTGGCAGAGAGCTTTCTTTAATTAACATCCAGAGTCTCC 11881
QY 218 ThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237
DB 11882 ACCATGAGTGGTGGGAAACCATTAAGACATTTGGCCAAACAAGAAATGATTTCATTTT 11941
QY 238 AspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGln 257
DB 11942 GACATCAAACTCCCTTGATCATGCTTAAATAATGATTTTCTTTTACTGACATTCAA 12001
QY 258 HisValThrAspLysIleIleMetLysLysGlyLysGln----- 271
DB 12002 CACGTTACTGATTAATAATAATTAAGAAAAAGAAAGAGATATTTCTAATTAAGT 12061
QY 272 -----G1 272
DB 12062 AAAGTTGAACAGATAAATATTCCTGATGTCCTTAAAAATAGCTGCTCATTTCTGACG 12121
QY 272 YlleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr 292
DB 12122 TATTGAGAAAGGGCAAAATCCTGCTTCACACACAGACATTTGCTGCCACAGTACAGAC 12181
QY 292 PheAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAs 312
DB 12182 AATGGCATGTCATCTTGTGTAAGAAACACATGGGCTATTTCTGTTTGTAAAGAGAGA 12241
QY 312 PheLeuProGlnAsnAsnAlaValLeu----- 321
DB 12242 CTGTGTTACTCAAAATATATGACGTACTGTGATTTTATCTAATTTATATGTAATGTA 12301
QY 321 ----- 321
DB 12302 CACTTGCATATGTTACTTTTCCAAAGACCTTGACCTTGTTAGGATGACAGAT 12361
QY 321 ----- 321
DB 12362 CTTTATGCTTATGCTAGCCCTGACAGTATGAATTAATGACAGATGAGAAAGCTAACAG 12421
QY 321 ----- 321
DB 12422 CCATTTCTTGTACTAGTTGTAGCTTTATGGAACAGCTGTATAGCTTCTATGCACTA 12481
QY 321 ----- 321
DB 12482 AGCTAATTTTGCATCTTCTGTGATTTTAAAGAGGCTTACAAATAAAGAAATTAAT 12541
QY 321 ----- 321
DB 12542 GCAGTAACTGATATCATTTTATGAATAATAGTGATTTCTTCATCTTTGATGA 12601
QY 321 ----- 321
DB 12602 ATCCCTTGTGTTGTTTATTAATAGCAGATCAATTAATAGCAGTGGAGGTGTAT 12661
QY 321 ----- 321
DB 12662 TCCAACTTGTGACATTAATGTTGATTAAGTTCTGATATATCATTAATTTGACAGCC 12721

```

QY      321 ----- 321
Db      12722 AAAATCCCTTAATGTCCTTAAGAGCCTGACAAACATCCGTTTAACTGATCTTAA 12781
QY      321 ----- 321
Db      12782 CTTTATTCATTAAAAATTAATACTAAAGTGGAAAAATGTTTAAAGTAGTAATTCAT 12841
QY      321 ----- 321
Db      12842 AGATGGAATTTACATGATATCAAGAATAATTTTTCAGAGTTATGATGTAATAATGCA 12901
QY      321 ----- 321
Db      12902 CAATAATATAAAAAATTCAGAGGTCTAAAAATAGTACATATGTAATTAATTAATAATA 12961
QY      321 ----- 321
Db      12962 AATATTTAGATGAAGGTTGGAAGAAATATACAAAAATGCTAGTAATGTTGATGCTA 13021
QY      321 ----- 321
Db      13022 TTGAAATTAATTAATTTTCTTCCAAATTTTATTAACATAGATATGTCATCTGCC 13081
QY      321 ----- 321
Db      13082 CATTACCACCTCAAAAATGGATAGTTTATTTATGTTAATGCTGATATTTTCTCCAG 13141
QY      321 ----- 321
Db      13142 TTTAATTAGCAGCTGTCATATCCATATATGATAGTATTTTGTGTTTCTCAATTCCT 13201
QY      322 ----- 340
Db      13202 TCAGGTGCTGTCGTGTCGCAAGTACCTTCTATATCCGACAGCTTCGGAATTTT 13261
QY      340 ----- 360
Db      13262 AACCAAGGCAACACAGTCCACTTTGTTGTCTCTCCACAGCTATGACATGATTAATGG 13321
QY      360 ----- 365
Db      13322 CATATGATGTCATGTTAAGCAGACAGATATACGTCCTCACTAATCTATGTAATAT 13381
QY      365 ----- 365
Db      13382 TAATTGCCATTATTAATCACTAAGCCTTCTTCCAGATCTTGGAGCTATGATTTAT 13441
QY      366 ----- 379
Db      13442 TTTAATGCTTCTTATTAAGAAATGATTTGAAAGACTACGTGCTGGCTTGGCAATTTAC 13501
QY      379 ----- 389
Db      13502 ATGACATAGAAAGCATCCGCTATGACCAAAAG 13533

```

```

/ LENGTH: 94750
/ TYPE: DNA
/ ORGANISM: M. catarrhalis
/ FEATURES:
/ NAME/KEY: misc_feature
/ OTHER INFORMATION: incyte template ID No. 6632636 38
/ PUBLICATION INFORMATION:
US-09-596-002-38

Alignment Scores:
Pred. No.: 2.6e-50 Length: 94750
Score: 502.00 Matches: 129
Percent Similarity: 50.81% Conservative: 59
Best Local Similarity: 34.86% Identities: 149
Query Match: 23.62% Indels: 34
DB: 4 Gaps: 9

US-10-649-273-2 (1-414) x US-09-596-002-38 (1-94750)
QY      39 ----- 57
Db      23873 GATTGGAGTTGGAACATCTGTGATGAAACAGGCTTGCCATCTATGATAGTACGAAG 23932
QY      58 ----- 73
Db      23933 AATGGCGTCGTGGCGGCTGCTCAGCTTTGTATAGCCAAATTAATCACTACATGCC 23992
QY      74 ----- 93
Db      23993 ACCTATGGCGGTGTGTCGTCAGCTTGCAGCTCCAGACCAATTCGTAAAGCTTGGCCG 24052
QY      94 ----- 113
Db      24053 TTATTTAATGAGCTGTGATGACGGAATATACCAAAATCCGACATGATAGCGTGGCG 24112
QY      114 ----- 133
Db      24113 TATACCAAGGCGCAGGCTGATTTGGCGGTGATGACAGCGGCACTTATTTGGCGCGAG 24172
QY      134 ----- 153
Db      24173 CTGGCGTATGGGCTGGCGCTGCAGCGGTGGGCGCATATATGAGGACATCTGTTA 24232
QY      154 ----- 171
Db      24233 GCACCGCTATTTGGCCAGTGAATACCCCCCAAGCTTTCGTTGTGCTGCTGCGTGTG 24292
QY      172 ----- 191
Db      24293 GCGGCTCATACATGCTGTGCTGCCATGCTGTGCGGTGATACATATTTGGCGCGAG 24352
QY      192 ----- 211
Db      24353 TCTATCGATGATGCGGTGGTGAATGCTTGTAAACCGCAAAATGCTCAAAATG--- 24409
QY      212 ----- 231
Db      24410 CCTATTCCT-----GGTGGCCCAAAATATGAAAAATTAACCAAAAAC 24451
QY      232 ----- 251
Db      24452 GGCACCCACACAGCCTATGAGCTGCCAAGACCCATGACAGAT---AAAGGCGTGATTTT 24508
QY      252 ----- 271
Db      24509 TCGTTCAAGGCGATGAAACCGCCATTCATATCTCATCAAGAACACCAACCGCCAA 24568
QY      272 ----- 291
Db      24569 AGCGAGCCC-----GCCACACGACGACATGCGCGCAAGCTTGTGATAT 24613
QY      292 ----- 311
Db      292 ThMeTAlaCybHleuValIybaRgThrhIaIdgaIaIleuPheCybYsGInaDg 311

```

```

RESULT 6
US-09-596-002-38
/ Sequence 38, Application US/09596002
/ Patent No. 6632636
/ GENERAL INFORMATION:
/ APPLICANT: Lagace, Robert, E.
/ APPLICANT: Patterson, Chandra
/ APPLICANT: Berg, Kim, L.
/ TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
/ FILE REFERENCE: PM-0008-4 US
/ CURRENT APPLICATION NUMBER: US/09/596, 002
/ PRIOR FILING DATE: 2000-06-16
/ PRIOR APPLICATION NUMBER: 60/140,121
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: PERL Program
/ SEQ ID NO 38

```



```

Db      24614 GCGGTGTGATCTTGTGTCAAAAATGCACCAAGCACTACAGATGACAGGCAATTCCG 24673
Qy      312 AspleuleuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspHe 331
Db      24674 CAG-----CTGGTGTGCGCAGGGGGCGTCTCTGCAATGCG 24709
Qy      332 TyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysPro 351
Db      24710 ATGCTAGCCGCCACCTGACCGCAGACGCTCCGCCAATGATGCGGTGTGATGATGCG 24769
Qy      352 ProProArgLeuGlyGlyThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeu 371
Db      24770 CCGACCGAGCTATGACGGAATATGATGCGATGCGCTTATGCGCTTTGTGCGGCTC 24829
Qy      372 ArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyTrpGluProLysCysPro 391
Db      24830 AGCTGTGACAGCTGC-----GATACCTTGGC-GGTTCGCTGTATTC-----CCG 24873
Qy      392 LeuGlyValAspIleSerLysGluValGly 401
Db      24874 ATGGGATATGACGACGCTTGGCGTATCGGC 24903

RESULT 7
US-09-540-236-806
; Sequence 806, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 806
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: M.catarhalis
US-09-540-236-806

Alignment Scores:
Pred. No.:      1,236-53      Length:      1053
Score:          501.00      Matches:      126
Percent Similarity: 49.86%      Conservative: 56
Best Local Similarity: 34.52%      Mismatches: 139
Query Match:      23.58%      Indels:      44
DB:              Gaps:      8

US-10-649-273-2 (1-414) x US-09-540-236-806 (1-1053)
Qy      39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThr--- 57
Db      16 GTATTGGGATTGGAGACATCTTGATGAAACAGGCGTGCATCATGATGATGACGATG 75
Qy      58 -----GlyAsnValLeuGlyLysAlaIleHisSerGlnThrGluValHisLeuLys 74
Db      76 AATGGCAGGGCGGCGTGTCTCTCAAGTTTATATACCAATTAATCATCATCCACC 135
Qy      75 ThrGlyGlyIleValProProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgLe 94
Db      136 TATGGCGGTGTGCTGCTGACCTTGCAGTCGACGACCATTCGTAAGCTTGTGCCGTTA 195
Qy      95 ValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThr 114
Db      196 TTTAATAGAGATTGATCAGGCAAAATATCACAAATCGACATTCATGATGCGGTGAT 255
Qy      115 ThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu 134
Db      256 ACCCAAGGCCCGCGGCTGATTTGGCGGATGACAGGGGCGCATTTTGGCGCGACGCTG 315
Qy      135 ValGlyGlnLeuLysLysProPheIleHisIleMetGluAlaHisIleAlaLeuThr 154
Db      316 GCGTATGGCGGTGCGGTGCGACCGGTGGGATGATCATATGAGAGGACATCTTTAGCA 375

```

```

Qy      155 IleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSerGly 172
Db      376 CCGCTATTGTGCGCAGTATGACCCCAAGCTTTCGTTTGTGTGCTGTGTGTGCGGC 435
Qy      173 GlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSer 192
Db      436 GGTATACCATCATGTGTGTGCGTGCATGCGTGTGCGGTGATCATGATATTTGGCGAGTCC 495
Qy      193 LeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLys 212
Db      496 ATCGATGATGCGGTGGCGCAATGCTTGATTAACCGCAAAATGTCTCAAACTG---CCC 552
Qy      213 HisProGluCysSerThrMetSerGlyLysValAlaIleGluHisLeuAlaLysGlnGly 232
Db      553 TATCTCT-----GTTGTCCAAATATGCAAAATTAAGCCAAACGCG 594
Qy      233 AsnArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSer 252
Db      595 AACCCACAGCCCTATGAGCTGCCAAGACGATGACGAT---AAGGCGTGTATTTTTCG 651
Qy      253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluGlyGly 272
Db      652 TTCAGTGGCATGAAAACCCGCAATTCATATCTCATGAAGACACAGCCCAACGCCAAAGC 711
Qy      273 IleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr 292
Db      712 GACCT-----GCCACACAGCAGACATGCGCCGCAAGCTTGTGATGCG 756
Qy      293 MetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAsp 312
Db      757 GTGTGTGATACTTGTGTCAAAATATGACCAAGGCACTACAGATGACAGGCAATGCCAG 816
Qy      313 LeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyr 332
Db      817 -----CTGGTGTGCGCAGGGCGTCTCTGCAATGACAGC 852
Qy      333 IleArgArgAlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCysProPro 352
Db      853 CTACCGCGCACCTTACCGCAGACGCTGCGCAATGATGCGGTGTGATGATGACGCGCCG 912
Qy      353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArg 372
Db      913 ACCGAGCTATGACGCGATATGCTGCGATGATCGCTTACGCTGCTTTGTGCGGCTAAGC 972
Qy      373 AlaGly-----Leu 375
Db      973 CGTGGCAGTCCGAGTATGCGGTGCTGTATTCGCCGATGGATATGACGATGCTT 1032

RESULT 8
US-09-252-991A-801
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

```


US-09-252-991A-801

Alignment Scores:

Pred. No.: 2,43e-53 Length: 1206
 Score: 499.50 Matches: 133
 Percent Similarity: 51.68% Conservative: 67
 Best Local Similarity: 34.37% Mismatches: 158
 Query Match: 23.51% Indels: 29
 DB: 4 Gaps: 10

US-10-649-273-2 (1-414) x US-09-252-991A-801 (1-1206)

```

QY 13 LysProSerLysArgLysValTyrGluPheLeuArgSerPheAsnPheHisProGlyThr 32
DB 91 AAAGCCTCCAGTCAAAAACCTGCTGGGCGCAAAAGCCGCGATTATCATCGGGCTT 150
QY 33 LeuPheLeuHisLysIle-----ValLeuGlyTyrLeuThrSerCys 46
DB 151 CGTATTGTCAGAGCTTAAGGCTCAAGCCCATGCGGCTGCTGGACTGGAAACGTCTTC 210
QY 47 AspAspThrAlaAlaValValValAspGluThrGlyAsnValLeuGlyAlaIleHis 66
DB 211 GACGAAACCGGCGCTGCGCTTATACGACGCAACCGCGCTGCTGCGCGCTGCTTC 270
QY 67 SerGlnThrGluValHisLeuLysThrGlyGlyIleValProProAlaIleGlnLeu 86
DB 271 AGTCAAGATCAACCTCCATCGGCTCAAGCGGCGCTGCTGCGGCTGCTGCGCGAC 330
QY 87 HisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSerAlaSerGlyValSerPro 106
DB 331 CACGTCAGAGCCGATCTCGCTGATTCGCCGCGCGAGTCTGACAGATCGCGCTGACAGCG 390
QY 107 SerLeuLeuSerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyVal 126
DB 391 GCGATATACGACGCGATGCGCTATACCGCGGCTGCTGCGGCTGCTGCTGCTGCTG 450
QY 127 GlyLeuSerPheSerLeuGlnLeuValGlnLeuLysLysProPheIleProIleHis 146
DB 451 GGGGCTTCTTGTCCCAAGCGATGCGCTTCCGCTGGGCGCTGCGCGGCTGCGCTGAC 510
QY 147 HisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPhe 165
DB 511 CACATGGAAGGCAACCTGCTGGCGCGGAGTGTGAAGACACCAACCGCGGTTCCGCTTC 570
QY 166 LeuValLeuLeuLysSerGlyValHisCysLeuLeuAlaLeuValGlnGlyValSerAsp 185
DB 571 GTCCGCTTGTGGGCTTCCGCGGCTGACACCCAGTGTGGTGGCGGCTGATCGCGCGC 630
QY 186 PheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAla 205
DB 631 TACCAATTGCTTGGGAATCGGTGAGACATGCGCGGCGAAAGCTTTCACAAAGACCGC 690
QY 206 ArgArgLeuSerLeuLysHisProGlyLysSerThrMetSerGlyLysValAlaIle 225
DB 691 AAGCTGATCGGCTG---GGCTATCCC-----GGTGTCCGGGAATC 729
QY 226 GluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHis 245
DB 730 GCCCGCTGGCGGAGCGCGGCACTCTGCGCGCTTGTGTTCCCGCGCGGATGACCGAT 789
QY 246 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleLeuMet 265
DB 790 CGCCCGGCGCTGAGACTTCAAGCTTCAAGCGGCTCAAGACCTTATACCTTGAAAC--ACTGG 846
QY 266 LysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIle 285
DB 847 CAGCGTTCCGTCGAGCGCGGACGACACGACGACG---ACCGCTGCGACATC 897
QY 286 AlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIle 305
DB 898 GCCCTGGGCTTCAAGACCGCGGTGTGACGACCTGCTGATCAAGTCCGCGCGCTTG 957
QY 306 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGly 325
  
```

```

DB 958 -----AACGACCGCGGCTG-----AAGAAC-----CTGGGATCGCGGC 993
QY 326 GlyValAlaSerAsnPheThrIleLeuArgAlaLeuGlnIleLeuThrAsnAlaThrGln 345
DB 994 GGTGTCAAGGCCAACGAGCGGCTGCGACGCGCTGGAAGAGTGTCCGCGAATGAC 1053
QY 346 CysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 365
DB 1054 GGGCAGGTGTCTACGCGCGCGCTTCTGACCGACATAGCGCGATGATCGCTTAC 1113
QY 366 AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArg 385
DB 1114 GCGGCTGCGACGCGCTGCTGCGCGC-----CAGCATACCGCGCGGATCAGC 1164
QY 386 TyrGluProLysCysProLeu 392
DB 1165 GTCCAGCCGCGCTGCGCGATG 1185
  
```

RESULT 9

```

US-09-252-991A-884/C
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884
  
```

Alignment Scores:
 Pred. No.: 1.52e-52 Length: 1059
 Score: 492.50 Matches: 128
 Percent Similarity: 53.24% Conservative: 61
 Best Local Similarity: 36.06% Mismatches: 143
 Query Match: 23.18% Indels: 23
 DB: 4 Gaps: 9

US-10-649-273-2 (1-414) x US-09-252-991A-884 (1-1059)

```

QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGly 58
DB 1038 GTGCTGGAGCTGAAACGTCTGCGCAAGAACCGGCGTGCCTTATACGACAGCAACGC 979
QY 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIle 78
DB 978 GGCCTGCTGCCGACGCGCTGTTCATGCAATCGACTCCATCGCTTACGAGCGGCTC 919
QY 79 ValProProAlaAlaGlnIleLeuHisAspGluAsnIleGlnArgIleValGlnGluAla 98
DB 918 GTGCCGAGCTGCGCTTCCGCGACACGTCMAAGCGATGCTGCGCTGATCGCGCAGTG 859
QY 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118
DB 858 CTCGAGAGATCCGGGCTGACGCGCGGATATGACGCGATGCTTATACGCGCGTCCC 799
QY 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138
DB 798 GGCCTGTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 739
QY 139 LysLysProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThr 158
  
```

```

Db      738 GAGCGCCGCGCGCGCGCGCGCGCAACATGAGAGGCACTGTCGCGCCGATGTCGAA 679
Qy      159 Aenlyr---ValGluPharProPheLeuValLeuLeuLeuSerGlyGlyHisCysLeuLeu 177
      678 GAGCAGCCACCGCGGTTCCCGTTCGTCGCGCTTGGTTCCGCGCGTACACCGCATG 619
Db      178 AlaleuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197
Qy      618 GTGCGGAGTGAAGGATGCGCGCGCTACAGATTGCTTGGCAATCGGTGACGATCGCGC 559
Db      198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSer 217
Qy      558 GCGCAAGCGCTTCGACAAAGCCGCAAGCTGATCGGCTG---GCGATATCC----- 511
Db      218 ThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237
Qy      510 -----GGTGGTCCGGAATTCGCCCGCGCTGGGAGAGCGGCACTTCGCGCGCTTC 460
Db      238 AspIleLysProProLeuHisHisValLysAsnCyAspPheSerPheThrGlyLeuGln 257
Qy      459 GTGTTCCCGCGCGCGATGACCGATCGCGCGCGCTGACATTCAAGCTTCAGCGGCTCAAG 400
Db      258 HisValThrAspLysIleIleMetLysGlyLysGlyGlnGlyIleGluLysGlyGln 277
Qy      399 ACCTTTACCTTGAC---ACCTGCGACGCTTCGTCGAGCGCGGACGACAGCGAGCAG 343
Db      278 IleuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeu 297
Qy      342 -----ACCGCTGCGACATCGCGCTGCGCTTCGACAGCGCGGTGTGACAGCCTG 292
Db      298 ValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGluAsn 317
Qy      291 CTGATCAAGTCCGCGCGCTG---AGCAGACCGGCTG-----AAGAAC 247
Db      318 AsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgArgAlaLeu 337
Qy      246 -----CTGGTATCGCGCGCGGTGTGACGCGCAACAGCGCGCTGCGAGCGGCTG 196
Db      338 GluIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCyAspProProArgLeuCyThr 357
Qy      195 GAAAGAAGATCTCGCGGAATGAGGGGAGGTGTTCTACGCGCGCGCTTCTGCAAC 136
Db      358 AspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIle 377
Qy      135 GACATAGCGCGGATATGCTTACCGCGCTGCGCGCTGCTGCTGCGCGC----- 85
Db      378 LeuHisAspIleGlnGlyIleArgGlyLeuProLysCyAspProLeu 392
Qy      84 CAGCATGACGCGCGCGATGACGCTCAAGCGCGCTGCGCGATG 40

RESULT 10
US-09-543-681A-2341
; Sequence 2341, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: CARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2341
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2341

Alignment Scores:
Pred. No.: 2,97e-51      Length: 1074
Score: 482.50      Matches: 122

```

```

Percent Similarity: 49.58%      Conservative: 55
Best Local Similarity: 34.17%      Mismatches: 149
Query Match: 22.71%      Indels: 31
DB: 4      Gaps: 8

US-10-649-273-2 (1-414) x US-09-543-681A-2341 (1-1074)
Qy      39 ValLeuGlyIleGlnThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58
Db      58 GTTTTAGGTTTGAACATCTTTCGATGAAACCGGATGCGCAATTACGATTAAGCC 117
Qy      59 AsnValLeuGlyLysAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIle 78
Db      118 GGTCTGTAGCGAATCACTCTATACCCAAATAACTGCAACCGGCACTATGGTGGTGT 177
Qy      79 ValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAla 98
Db      178 GTTCCGACCTTCTGACGAGTCAATTCGCGAAACAGTGCACATGATCCAGCGCGC 237
Qy      99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118
Db      238 CTTAAAGAAACAATTAACTGCAAAAGATATGATGGGTGCTTATCTGACAGACCA 297
Qy      119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuGlyGlnLeu 138
Db      298 GGGCTGTAGGGGCAATTGCTGCGTGGGCGGACCATGTTGTTCTGCGGCTTGCATGG 357
Qy      139 LysLysProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThr 158
Db      358 GATGCTCTGATTCATTCACATTCACATATGAAAGGCGCATTTATTAGCGCGCATGAA 417
Qy      159 AenlyrVal---GluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeu 177
Db      418 GAGAAAGCCCAAGATTTCTTCTGCGGCTTACTGTCTCGGGGGGCGCATACAAATTA 477
Qy      178 AlaleuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197
Db      478 ATTAGTGTAACAGGATGCGGAATATACCTCTTAGTGATGATGATGATGATGATGATG 537
Qy      198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSer 217
Db      538 GGTGAAGCATTTGATTAACAGCCAGCATATTAGGCGCTT---CATATATCC----- 585
Qy      218 ThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyLysAsnArgPheHisPhe 237
Db      586 -----GGCGGCGCTGTTTATCAAAAATGCAACAAAGGTGTAGAGGACGTTT 636
Qy      238 AspIleLysProProLeuHisHisValLysAsnCyAspPheSerPheThrGlyLeuGln 257
Db      637 GTTTTCTCTGCTCCATGACAGACAGACCGGACCTGACTTTAGTTTCTCTGATTAATA 696
Qy      258 HisValThrAspLysIleIleMetLysGlyLysGlnGlyGlnGlyIleGluLysGlyGln 277
Db      697 ACCTTGGCGGCTATGATCTATGTCGTAAGACATGATTCAGAG----- 738
Qy      278 IleuSerSerAlaAlaAspIleAlaIleArgValGlnHisThrMetAlaCysHisLeu 297
Db      739 -----CAAACTCGACAGATATGCGCGCTGTTGAAGATCCGATGATATCTTGG 792
Qy      298 ValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGluAsn 317
Db      793 GCAATAAATGTCGTGACGA-----TTAGAGCAACA 825
Qy      318 Asn---AlaValLeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgAla 336
Db      826 GCGTTTAAACGCTTAGTATGCTGGGCGCGTAAAGCTCAACGCTTATCCGCGCAAA 885
Qy      337 LeuGlnIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCyAspProProArgLeuCy 356
Db      886 ATGCGGATGATATAGAACAACTCGAGGGGAAGGTGTTTATGCTGCGCGCTGATATGT 945
Qy      357 ThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGlyLeu 375

```

Db	946	ACCGATATAGTGCATATATTGCTTTGGCGGGAGNATCCGTTTAAAGATGTAACGAG	1005
Qy	376	GLYILLEUHSAPPILEGUIGLYILEARGYRGLUPOLYSCYSPROLEU	392
Db	1006	GGGCGCTTTA-----GGGGTGACAGTGAACCAACCGTGGCGCTTTA	1044
RESULT 11			
US-09-902-540-6612			
; Sequence 6612, Application US/09902540			
; Patent No. 6833447			
; GENERAL INFORMATION:			
; APPLICANT: Goldman, Barry S.			
; APPLICANT: Hinkle, Gregory J.			
; APPLICANT: Slater, Steven C.			
; APPLICANT: Wisgand, Roger C.			
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof			
; FILE REFERENCE: 38-10(115849)B			
; CURRENT APPLICATION NUMBER: US/09/902,540			
; PRIOR FILING DATE: 2001-07-10			
; PRIOR APPLICATION NUMBER: 60/217,883			
; PRIOR FILING DATE: 2000-07-10			
; NUMBER OF SEQ ID NOS: 16825			
; SEQ ID NO 6612			
; LENGTH: 996			
; TYPE: DNA			
; ORGANISM: Myxococcus xanthus			
US-09-902-540-6612			
Alignment Scores:			
Pred. No.: 3,01e-51 Length: 996			
Score: 482.00 Matches: 128			
Percent Similarity: 52.48% Conservative: 52			
Best Local Similarity: 37.32% Mismatches: 131			
Query Match: 22.68% Indels: 32			
DB: 4 Gaps: 9			
US-10-649-273-2 (1-414) x US-09-902-540-6612 (1-996)			
Qy	38	lIleValleuglYlIeGlUthSerCYsaPhaapThralaalaIaValaIaspGlUthR	57
Db	4	CTCGTCTCAGACTGAGAAACCTCGTGTATGAGACTGCCGCGCGTGTGAGAGGACGC	63
Qy	58	GLYasnaValleuglYgluaIaIaIeHisserGlnthrglUvalHIsleuYethrGlyGly	77
Db	64	CGCCGCGGCTGTGCAGTGTCTCCACGACGATGACATCCGCGGTGGGTGGG	123
Qy	78	lIleValProProlaIaIaIeGlnleuHIsarGluanaIeGlnarglIeValGlnGlu	97
Db	124	GTGTGTCCGAGAGCTGGCCAGCCGCAACATCTGCCAGGTGTGCCCTGTCTCAGCAG	183
Qy	98	AlaIeuSerAlaSerGlyValSerProSerApsleuSerAlaIaIaIaIaThrThrIleYs	117
Db	184	GGCGTGAAGCGGGCGAACAAGCGCTGACACGACGACATCATGCGGTACGTCGCGG	243
Qy	118	ProGlyLeuAlaIeuSerleuGlyValGlyIeuSerPheSerleuGlnleuValGlyGln	137
Db	244	CCCGAGCTCATCGGCGCGCTGTGTGTGAAGTGAAGTGGCCAGGAGGCTTGAAGCTTGG	303
Qy	138	LeuYlsPProPheIleProIleHIsHIsMetGluAlaHIsAlaIeUthRlIeaYrleu	157
Db	304	ACGGGCAAGCCCTTCTGTGTGGGCGCCCAACATCTGAGAGGCCACATCTGGCCATCGAG	363
Qy	158	ThraInYsVal---GluPheProPheleuValleuIleuIleSerGlyYHIsCYsleu	176
Db	364	TTGAGAGTGGCGCGGAGCGCGCGTTCCTTGTGGCTCGTTCGTTCCGGGGGACACGAC	423
Qy	177	LeuAlaIeuValGlnGlyValSerApsPheIeuLeuIeuGlyYlsSerleuApsIleAla	196
Db	424	CTTACAGAGTGTGACGGCTTACGCGACATACGCGCTGTGTGGCAGACACCGACACGCG	483
Qy	197	ProGlyApsMetleuApsYlsValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa	216
Db	484	CGCGGCGAGGATATGATGAAGACCGCTGTGATCTTGGGCTG---CGATATCG-----	534

```

Oy 217 SerThrmEserGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsp----- 233
Db 535 -----GGTGGGACGCCCATGACAGTGGCGAGCGAGGGAAACCGGAGGCC 582
Oy 234 ---ArgPheHisPheAspIleLysProProLeuHisIleAlaLysAsnCyAspPheSer 252
Db 583 ATCGGCGTTT-----CCCGCGCGCGTCCGCGGCGACAACTTCGACGTCTCC 627
Oy 253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGlnLysGluGluGly 272
Db 628 TTCCTCCGGGTTTGAAG-----ACCGCGGCTGTCGACCACTGCGAAGACACGCG 675
Oy 273 IleGluLysGlnGlnIleLeuSerSerAlaAlaAspIleAlaThrValGlnHisThr 292
Db 676 GTGCGCGAGGGGAGCGCGCGCTG-----GCGGATTTGCGCGGTTCTTCCAGAGGCC 726
Oy 293 MetAlaCyHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAsp 312
Db 727 GTGCGCGAGCGTGTGCGAAGAG-----CTGGTGCGCGCGCGCGCGCG 771
Oy 313 LeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyr 332
Db 772 TTG-----GGCCACAGCAAGTTGTGTGTGCGCGCGCGCTCGCGGAACCTCGCG 822
Oy 333 IleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCySProPro 352
Db 823 CTGCGCGGACCTGTGTACGCGCGCGAGCGGAGCGCGGGGTGTGAACATGTTCTCGCCCG 882
Oy 353 ProArgLeuCySThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluArgLeuArg 372
Db 883 GTGCGGCTGTGTACAGCAATGTGGCGCATGATTGCGGTGCGGGGTATGAGCGTACCGC 942
Oy 373 AlaGlyLeu 375
Db 943 CGCGGCGCTG 951

RESULT 12
US-09-902-540-503/C
; Sequence 503; Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217, 883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 503
; LENGTH: 2582
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-503

Alignment Scores:
Pred. No.: 1,62e-50 Length: 2582
Score: 482.00 Matches: 128
Percent Similarity: 52.48% Conservative: 52
Best Local Similarity: 37.32% Mismatches: 131
Query Match: 22.68% Indels: 32
DB: 4 Gaps: 9

US-10-649-273-2 (1-414) x US-09-902-540-503 (1-2582)
Oy 38 IleValLeuGlyIleGluThrSerCyAspAspThrAlaAlaValAlaAspGluThr 57
Db 1525 CTGCTCTTAGACATCGAAACCTGCTGTGTATGACACTCCCGCGCGTCTGTGAGAGACGCC 1466

```

QY 58 GlyanValleuglylualalIehiserglnthrgluvalhlsleuylthrglygly 77
 Db 1465 CGCCGCGGCTGTGCGATGTCCTCCACGAGGTGACATCCACGCGGGTGG 1406
 QY 78 lIeValProPfoalalaglnghleuhsargluamnlleglnarglIeValglnlu 97
 Db 1405 GTGGTGTCCGAGCTGGCCAGCCGACACATCTGTCAGTGTGCTGCTCCACGAG 1346
 QY 98 AlaleuseralaserglyValserProserAspleuseralIlealathrhrlleys 117
 Db 1345 GCGCTGACGCGGCGCAACAGAGCTCCAGACGAGCTCATCTGCTCCAGCTCGGC 1286
 QY 118 ProglyleualaleuserleuglyValglyleuserPheuserleuclnleuValglyln 137
 Db 1285 CCGGAGCTCATCGGCGGCTGTGTGGAGTGCAGTGGCCAGGCGCTTGAGCCTGGGG 1226
 QY 138 LeuylslyPProPheleProlelshlsmetglualahlsalaleuthrIleargleu 157
 Db 1225 ACGGCGAAGCCCTTCTGTGGGCGCAACACCTGAGAGGCGCACTGCTGGCCATCCGGCTG 1166
 QY 158 ThranslyVal----GluPheProPheleuValleuLeuIleSerglylyhsybleu 176
 Db 1165 TTGAGAGTGGCGCGGAGCCCGCTTCTGTGGGCTGTGCTTCCGCGGCGACACGAC 1106
 QY 177 LeuAlaleuValglnlyValserAspleleuLeuGlylyserleuapllea 196
 Db 1105 CTTCACAGAGTGCAGGCTTACGGGAGTACCGGCTGTGTGGAGCAGCGCACAGCGC 1046
 QY 197 ProglyAspleleuAspleValalalargargleuserleuIlelyshlsProglucy 216
 Db 1045 GCCGCGAGGATATGACAGACGCTCGCATCTCGGCTG---CCGATCCG----- 995
 QY 217 SerthmetserglylyValalaleuhsleuAlalysegllyAsn----- 233
 Db 994 -----GGTGGCGAGCCCATCCACCATGTCGCGCAGGAGGAAACCGGAGGCC 947
 QY 234 ---ArgPhehsPheaplleleuProProleuhsIshlsAlalyAsnCyAspPheSer 252
 Db 946 ATCCGCTTC-----CCGCGCGGCTGCGCGGCGACAACTTCAGCTGTCC 902
 QY 253 PheThrglyleuglnhlsvalThrasplyslelshlsleuylsleuglnleugly 272
 Db 901 TTCTCCGGGTTGAAG-----ACGGCGGTGCTGCACACAGTGCAGAGACGCGC 854
 QY 273 lIeuglyseglylnIleuserSereralalalapllealaleThValglnhlsThr 292
 Db 853 GTGCGGAGGCGGCGCGCTG-----GCGGATTTGTGCGGCTTCCAGGAGGCGC 803
 QY 293 MetAlaCyshlsleuVallyasargThrhlsarglalleuPheCylyseglInargAsp 312
 Db 802 GTGCGGAGCTGTCTGTGAGAGAG-----CTGTGGCGCGCGCGCGCGCG 758
 QY 313 LeuLeuProGlnAsnAlalavalIleuValalaserglylyValalaserAspPheTy 332
 Db 757 TTG-----GCCACAGACAGTGTGCTGTGCGCGCGGCTGCGCGGACTCGCGG 707
 QY 333 lIearglrgalaleuglnIleuThrasnAlathrglnCyethrleuCyAspPro 352
 Db 706 CTGCGGAGCTGTGTGAGCGCGAGCGGAGGCGGGGTTGAACATGTTCTGCCCCG 647
 QY 353 ProargleuCyethrAspAsnglylylshmetllealatrPhanglylyllegluargleuarg 372
 Db 646 GTGCGGCTGTGCAGGACAAATGGCGCCATGTGCGGTGGCGGGTATAGAGCGTACGCG 587
 QY 373 Alaglyleu 375
 Db 586 CGCGGCTG 578

? APPLICANT: Mellors, Alan
 ? APPLICANT: Lo, Reggie Y.C.
 ? APPLICANT: Abdullah, Khalid M.
 ? TITLE OF INVENTION: Pasteurella Haemolytica
 ? TITLE OF INVENTION: Glycoprotease
 ? TITLE OF INVENTION: Gene and the Purified Enzyme
 ? NUMBER OF SEQUENCES: 3
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSER: Bell, Selzer, Park & Gibson, P.A.
 ? STREET: 1211 East Morehead Street,
 ? CITY: Charlotte
 ? STATE: No. 554312ch Carolina
 ? COUNTRY: United States
 ? ZIP: 28214
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent In Release #1.0, Version #1.25
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/087,797
 ? FILING DATE: 14-JUL-1993
 ? CLASSIFICATION: 435
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Layton, Jr., Samuel G.
 ? REGISTRATION NUMBER: 22807
 ? REFERENCE/DOCKET NUMBER: 3374-80
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 704 334 2014
 ? TELEFAX: 704 377 1561
 ? INFORMATION FOR SEQ ID NO: 1:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 1315 base pairs
 ? TYPE: nucleic acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: cDNA
 ? HYPOTHEICAL: NO
 ? ANTI-SENSE: NO
 ? FEATURE:
 ? NAME/KEY: CDS
 ? LOCATION: 141..1115
 ? US-08-087-797-1
 ?
 ? Alignment Scores:
 ? Pred. No.: 1.19e-50 Length: 1315
 ? Score: 479.00 Matches: 128
 ? Percent Similarity: 48.63% Conservative: 50
 ? Best Local Similarity: 34.97% Mismatches: 157
 ? Query Match: 22.54% Indels: 31
 ? DB: 1 Gaps: 8
 ? US-10-649-273-2 (1-414) x US-08-087-797-1 (1-1315)
 QY 9 GlyValPhePheleuProserlyAsarglyValIyrgluPheleuArgserPheanPhe 28
 Db 89 GGGGCGCTTCTGCGCCCT-----TTGGTTTCTTAACCTTAATTTGACTTC 133
 QY 29 HisProglyThrleuPheleuhslyslle-ValleuglylyllegluThserCyAspAs 48
 Db 134 TCC-----AACTAGCGAAATTTAGGTATTAACCTCTTGATGA 175
 QY 48 pThralalalavalalaspIlyThrglyasnValleuglylualalIehisergl 68
 Db 176 AACCGGTGTGCATTATGATGAGACAAAGGCTTATGGCAACAGCTTATAGCCA 235
 QY 68 nThrgluValhlsleuylsThrglylylIeValProPfoalalaglnghleuhsar 88
 Db 236 AATTGATATGACAGCGCGCTTACGATGATGATGATGATGATGATGATGATGATGAT 295
 QY 88 ggluamnlleglnarglIeValglnghleuhsleuValalaserglylyValserProserAs 108
 Db 296 CCGTAAACGTTGCCATTAATTAAGAGGCGCAATCTGCAACCTCGGA 355

RESULT 13
 US-08-087-797-1
 ; Sequence 1, Application US/08087797
 ; Patent No. 554312
 ; GENERAL INFORMATION:


```

Dh 552728 GCGGAGCGCTTGATGATAAAGACGAAATAATCACTTGACCTA---GATTATGCA----- 552681
Oy 218 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyValAsnArgPheHisIlePhe 237
Dh 552680 -----GGTGGGCGCGCACCTTCTCGTTTAGCGGAAAAAGGTACGCGCAATCGTTTC 552630
Oy 228 AspIleLeuProIleLeuHisHisAlaValAsnGlyAspPheSerPheThrGlyLeuGln 257
Dh 552629 ACATTTCACGTCGCAATGACACAGATCGTCAGCGCGCTTGATTTAGTTAGTTTCTGGTTTAAAA 552570
Oy 258 HisValThrAspIleIleIleMetLeuGlyValGlyGluGluGly-----IleGluIys 275
Dh 552569 ACATTGGCGCGCAAAATACAGTTATATCAAGCAATTTAAAAAGAGGCGCAACTGATAGAGCAA 552510
Oy 276 GlyIleIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 295
Dh 552509 -----ACTAAACACAGATATGCTTATGCTTCCAGATCGCGGTGAGAT 552465
Oy 296 HisIleValIysArgThrHisArgAlaIleLeuPheCysIleGlnIleArgAspLeuPro 315
Dh 552464 ACTCTTGCC-----ATTAATATGTAAG---CGTGCATTGAAAGAA 552429
Oy 316 GlnAsnAsnAlaValIleValAlaSerGlyGlyValAlaSerAspPheThrIleArgArg 335
Dh 552428 ACAGCGCTATAAACGTTTACTGATTCGCGGAGGGGTGAGCGCAATTAATAAACTCCAGAA 552369
Oy 336 AlaIleGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysAspProProArgLeu 355
Dh 552368 ACGCTTGCGCACCTTATGCAAAATTTAGGTGCGCAAGTGTTTTATTCCTCAACCTCAATT 552309
Oy 356 CysThrAspAsnGlyIleIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGly 374
Dh 552308 TGATACAGATAAATGTGTGCGATGATGCTTACACAGAGTTTTTACGTTTAAAAACAAGGT 552252

RESULT 15
US-09-6643-990A-1/c
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS V6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-AUG-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PBI86PICI

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 1,21e-44 Length: 1830121
Score: 475.50 Matches: 120
Percent Similarity: 50.74% Conservative: 52
Best Local Similarity: 35.40% Mismatches: 144
Query Match: 22.38% Indels: 23
DB: Gaps: 7

US-10-649-273-2 (1-414) x US-09-643-990A-1 (1-1830121)
Qy 39 ValLeuGlyIleGluThrSerCysAspAerThrAlaAlaValValAspGluThrGly 58
Db 553208 ATCTTAGGCAATTGAACCTTCTCTGATGAACAGCGCGCGCATTTATGATGAAGAAAA 553149
Qy 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuYerThrGlyGlyLe 78
Db 553148 GGATTAATGCTATCACTTATTACTCAAAATGCGCTGCATGCAGATTAATGATGGTGATG 553089
Qy 79 ValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAla 98
Db 553088 GTTCCCTGAATTAGATCAACGATATATTCGAAAAACAGACCTCTTATTTAAAGGGCA 553029
Qy 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleYerPro 118
Db 553028 TTAAGAGAGCAATTTAAACCGCAGCATATTTGATGATTCCTTATACAGATGCGCTT 552969
Qy 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138
Db 552968 GGGCTGTGTGGCGGATCTTGTGGTGCTACAGATTCGACGTTCTTTACCTGATGCTTGG 552909
Qy 139 LysLysProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeu--- 157
Db 552908 AATGTTCTCGATGATGGTGATTCATATGGAAGGCAATTACTTCCGCCAATGCTGAT 552849
Qy 158 ThrAsnLysValGluPheProPheLeuValLeuIleSerGlyGlyHisCysLeuLeu 177
Db 552848 GACAAATCAACCGCACTTCTCTTGTGTCTCTGTGGTATGCGGATGGCCACACTCAATTA 552789
Qy 178 AlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197
Db 552788 GTGGTGTCATATGATGTAGAGAAATATGAAGATAGAGAGATCATATGATGATGCT 552729
Qy 198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSer 217
Db 552728 GGCAGAACCTTGTGATTAACACGCAAAATTAATCTGACATA---GATTATCCA----- 552681
Qy 218 ThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyLysAsnArgPheHisPhe 237
Db 552680 -----GGTGCGCGGCACTTCTCTGTTAGCGGAAAAAGTAGCGCAAAATCGTTTC 552630
Qy 238 AspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGln 257
Db 552629 ACATTTCCACGTCCTCAATGACAGATGTCGACAGCCCTTGATTTTAGTTTCTTCGTTAAAA 552570
Qy 258 HisValThrAspLysIleIleMetLysLysGluLysGlnGlyLys-----IleGluLys 275
Db 552569 ACATTTCCCGCCCAATTAACGTTATCAACCAATTAATAAAGCAGGGCGCAACTGATAGACAA 552510
Qy 276 GlyGlnIleLeuSerSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisPheMetAlaCys 295
Db 552509 -----ACTAAAGCAAGATTTATGCTTATGCTTTCCAAAGATGCGGATGCTGAT 552465

```

```
QY      296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315
      |||
Db      552464 ACTCTTGCC-----ATTAATGTAAAG---CGTCATTGAAGA 552429
      |||
QY      316 GlnAsnAsnAlaValLeuValAserGlyValAlaSerAsnPheTyrIleArgArg 335
      |||
Db      552428 ACAGGCTATAACGTTAGTAGTTCGGGAGGGGTGAGCGCAATTAATAAACTCCGAGAA 552369
      |||
QY      336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 355
      |||
Db      552368 ACGCTTGGCAGCTTAATGCAAAATTAGGTGGAGAGNGTTTATCTCAACCTCAATT 552309
      |||
QY      356 CysThrAspAsnGlyIleMetIleAlaTyrPAsnGlyIleGluArgLeuArgAlaGly 374
      |||
Db      552308 TGTACAGATAATGTGTGCGATGATTCCTTACACAGGTTTTCGTTTAAACAAAGGT 552252
```

Search completed: February 16, 2005, 21:50:50
Job time : 2493.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 18:03:13 ; Search time 828 Seconds

(without alignments)
2950.230 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125
Sequence: 1 MLITKTAGVFPKSKRKRV.....DISKEVGASIKVPLQKMEI 414

Scoring table:

BIOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+g2n.model -DB=xlh
-Q=/cgn2_1/USPTO.spool/US10649273/runat_14022005_114706_16497/app_query.fasta_1.1429
-DB=Published Applications NA -OPMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR.SCOR=spc -THR.MAX=100
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649273@cgn1_1.1053@runat_14022005_114706_16497
-NCPU=6 -ICPU=3 -NO.WMAP -LARGQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PTC_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PTC_NEW_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10F_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	2197	14	US-10-067-443-1
2	2125	100.0	2197	17	US-10-649-273-1
3	2125	100.0	2197	17	US-10-651-722-1
4	2090.5	98.4	1387	14	US-10-067-443-21
5	2090.5	98.4	1387	17	US-10-649-273-21
6	2090.5	98.4	1387	17	US-10-651-722-21
7	2088	98.3	1245	14	US-10-012-140-6
8	2088	98.3	1245	14	US-10-012-140-4
9	1944	91.5	2208	17	US-10-094-749-400
10	1944	91.5	2890	18	US-10-723-860-7447
11	1747	82.2	1416	17	US-10-120-948-177
12	1385	65.2	1526	14	US-10-067-443-23
13	1385	65.2	1526	17	US-10-649-273-23
14	1385	65.2	1526	17	US-10-651-722-23
15	1186.5	55.8	14364	14	US-10-067-443-20
16	1186.5	55.8	14364	17	US-10-649-273-20
17	1186.5	55.8	14364	17	US-10-651-722-20
18	682	32.1	1917	17	US-10-424-599-66417
19	578	27.2	1628	18	US-10-437-963-11249
20	531.5	25.0	4360	14	US-10-081-051-2
21	524	24.7	1146	17	US-10-282-122A-14674
22	502	23.6	1044	17	US-10-282-122A-26972
23	502	23.6	94750	17	US-10-672-787-38
24	494.5	23.3	1032	17	US-10-282-122A-31043
25	492.5	23.2	1026	9	US-09-815-242-7701
26	492.5	23.2	1026	9	US-10-282-122A-30016
27	488.5	23.0	1194	18	US-10-425-115-167432
28	484.5	22.8	1014	9	US-09-815-242-6207
29	484.5	22.8	1014	17	US-10-282-122A-20511
30	484.5	22.8	1014	17	US-10-282-122A-39301
31	483.5	22.8	1014	17	US-10-282-122A-41977
32	482.5	22.7	1020	17	US-10-282-122A-32254
33	481.5	22.7	1014	9	US-09-815-242-9682
34	475.5	22.4	1029	9	US-09-815-242-6946
35	475.5	22.4	1029	17	US-10-282-122A-22020
36	475.5	22.4	1830121	17	US-10-329-670-1
37	475.5	22.4	1830121	18	US-10-158-865-1
38	469	22.1	1023	17	US-10-282-122A-15870
39	468	22.0	371	18	US-10-430-201-3118
40	468	22.0	371	18	US-10-430-201-3119
41	467.5	22.0	1014	17	US-10-282-122A-21633
42	467.5	22.0	1007	17	US-10-282-122A-19220
43	465.5	21.9	1065	17	US-10-282-122A-29584
44	465	21.9	1023	17	US-10-282-122A-31809
45	462.5	21.8	936	17	US-10-282-122A-8315

ALIGNMENTS

RESULT 1
US-10-067-443-1
; Sequence 1, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 2,41e-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

QY 1 MetLeu11LeuThr1ySthrAlaG1yValPhePheLysProSeLysArgLysVal1Yr 20
Db 231 ATGCTAATCTTGACTAGACATGACGAGAGTCTTTTAAACCATCAAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLys11LeuValLeu 40
Db 291 GAATTTTAAAGAGTTTAAATTTTCATCTGGAAACATATTTCTTCATATAAATAGTATTTG 350
QY 41 Gly11LeuThrSerCysAspAspThrAlaAlaValValAspG1uThrG1yAsnVal1 60
Db 351 GGAATTTGAAACTTATGTTGATGATACACAGCTGCTGTGTGTGTGATGAACTGAAATGTG 410
QY 61 LeuG1yGluAla11LeuHisSerGlnThrG1yValHisLeuLysThrG1yL11LeuValPro 80
Db 411 TTGGAGAAAGCAATACATTTCCCAACTGAAAGTTCAATTTAAAAACAGGTGGATTTGCTCT 470
QY 81 ProAla11ArgGln1LeuHisArgGluAsn11LeuGlnArg11LeuAlaGln1uAlaLeuSer 100
Db 471 CCACAGCTCAACAGCTTCAACAGAAATATTTCAACAGATGTGTCAAGAAAGCTCTTCT 530
QY 101 AlaSerGlyValSerProSerAspLeuSerAla11LeuThrThr11LeuSerProGlyLeu 120
Db 531 GCCAGTGAAGTCTCTCCAGATGACCTCTCAGCAATGCACTTAAACCAATAAACAGGACTT 590
QY 121 AlaLeuSerLeuG1yValG1yLeuSerPheSerLeuGlnLeuValG1yGlnLeuLysLys 140
Db 591 GCTTTAAGCCCTGGAGGTGGGCTTATCATTTAGCTTACAGCTGTGTGTGAGACGTTAAAAAG 650
QY 141 ProPhe11LePro11LeHisHisMetGluAlaHisAlaLeuThr11LeuArgLeuThrAsnLys 160
Db 651 CCATTCATTTCCCATTCATCATATGAGGCTCATGCACTTACATTTAGGTTGACCAATATAA 710
QY 161 ValGluPheProPheLeuValLeuLeu11LeuSerGlyGlnHisCysLeuLeuAlaLeuVal1 180
Db 711 GTAGAAATTCCTTTTAAATTTTGAATTTCTGGAAGGTCACGTGCTGTGGCATTAGTT 770
QY 181 GlnGlyValSerAspPheLeuLeuG1yLysSerLeuAsp11Ala1ProG1yAspMet 200
Db 771 CAAGGAGTTTCAGANTTTCTGCTTCTTGAAAGTCTTTTGACATAGCACAGTGCATG 830
QY 201 LeuAspLysValAla1ArgArgLeuSerLeu11LeuHisProGlnCysSerThrMetSer 220
Db 831 CTTGACCAAGGTGGCAAGAACTTTCTTTAATTAACATCCAGATGTGCTCCACCATAGCT 890
QY 221 GlyGlyLysValAla11LeuHisLeuAlaLysGlnGlyAsnArgPheHisPheAsp11LeuLys 240
Db 891 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATAGATTTCATTTTACATCAAA 950
QY 241 ProProLeuHisHisAla11LeuAsnCysAspPheSerPheThrG1yLeuGlnHisVal1Thr 260
Db 951 CTTCCCTTGCATCAGCTTAAGAAATTTGTATTTTCTTTTACGTGACCTTCAACAGTTACT 1010
QY 261 AspLys11LeuMetLysGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280
Db 1011 GATTAATAATTAATGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1070
QY 281 SerAla11Asp11LeAla11ThrValGlnHisThrMetAlaCysHisLeuValLysArg 300

```

```

Db 1071 TCAGCAGCAGACATTCCTCCACAGTACAGCACACATGCAATGTCATCTTGGAAGA 1130
QY 301 ThrHisArgAla11LeuPheCysLysGlnArgAspLeuLeuProGlnHisAsnAlaVal1 320
Db 1131 ACACATCGGGCTATTCCTGTTTGTATAGCAGAGAGACTTGTACCTCAATAATATGACGTA 1190
QY 321 LeuValAlaSerG1yG1yValAlaSerAsnPheThr11ArgArgAlaLeuGln11Leu 340
Db 1191 CTGGTTGATCTGGTGTGTGTGCGACGTATCTTATATCGCAGAGCTCTGGAAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
Db 1251 ACMAAGCAACACAGTGCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1310
QY 361 11Met11Ala11TrpAsnGly11LeuArgLeuArgAlaG1yLeuGln11LeuHisAsp 380
Db 1311 ATATGATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1370
QY 381 11LeuGlnG1y11LeuArgThrGluProLysCysProLeuG1yValAsp11LeuSerLysGluVal 400
Db 1371 ATAGAAAGGATCGCTATGACCAAAATGTCTCTTGAGTACATATCAAAAGAAAGTT 1430
QY 401 GlyGluAlaSer11LeuSerValProGlnLeuLysMetGlu11Leu 414
Db 1431 GGAGAAAGCTTCCATATAAAGTACACAAATTAATAATGAGATA 1472

RESULT 2
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-649-273-1

Alignment Scores:
Pred. No.: 2,41e-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-1 (1-2197)

QY 1 MetLeu11LeuThr1ySthrAlaG1yValPhePheLysProSeLysArgLysVal1Yr 20
Db 231 ATGCTAATCTTGACTAGACATGACGAGAGTCTTTTAAACCATCAAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLys11LeuValLeu 40
Db 291 GAATTTTAAAGAGTTTAAATTTTCATCTGGAAACATATTTCTTCATATAAATAGTATTTG 350
QY 41 Gly11LeuThrSerCysAspAspThrAlaAlaValValAspG1uThrG1yAsnVal1 60

```

Db 351 GGAATTGAACTAGTGTGATGATACAGACGCTGCTGTGGATGAATGAAATGTG 410
Qy 61 LeuGIyGIuAlaIleHieSerGIInThrgIuValHieLeuYThrGIYIleValPro 80
Db 411 TTGGAGAGAAAGCAATACATTCCTCAAACTGAAGTTCAATTAAAAACAGTGGATGTCTCT 470
Qy 81 ProAlaIaGIInGInLeuHieSargGIuAenIleGIInargIleValGIInGIuAlaLeuSer 100
Db 471 CCAGCAGCTCAACAGCTTCACAGAGAAAATATTCAACGAATGATGACAGAACTCTTTCT 530
Qy 101 AlaseerGIYValSerProSerApleuSerAlaIleAlaThrIleYsProGIYleu 120
Db 531 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTGCACTACCAATAAACAGAGACTT 590
Qy 121 AlaleuSerLeuGIYValGIYleuSerPheSerLeuGIInleuValGIYInleuYsYs 140
Db 591 GCTTTAAGCTGGAGGTGGCTTATCATTTTACCTTACAGCTGTGAGACAGTTAAAG 650
Qy 141 ProPheIleProIleHieHieSmetGIuAlaHieAlaLeuThrIleArgLeuThraenYs 160
Db 651 CCATTTCATTCCTCATTCATCATATGAGGCTCATGCACTTACTATTAGTTGACCAATATA 710
Qy 161 ValGIuPheProPheLeuValleuLeuIleSerGIYGIYHieCysleuLeuAlaleuVal 180
Db 711 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACTGTCTGTGGCAATTAGTT 770
Qy 181 GIInGIYValSerApleuPheLeuLeuLeuGIYIYsSerLeuApleIleAlaProGIYAspMet 200
Db 771 CAAGAGATTCAGATTTTCTGCTCTTGAAAGCTTTTGACATGACACAGGTGACATG 830
Qy 201 LeuApleYsValAlaArgArgLeuSerLeuIleYsHieProGIYCysSerThrMetSer 220
Db 831 CTTCACAAAGGTGGCAGAGAACTTCTTTAATAAACATCCAGAGTCTCCACATGAGT 890
Qy 221 GIYGIYIYsValaIleGIuHieLeuAlaYsGIInGIYAsnArgPheHieApleApleIleYs 240
Db 891 GGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCATTTTACATCAAA 950
Qy 241 ProProleuHieHieSAlaYsAsnCysAplePheSerPheThrGIYleuGIInHieValThr 260
Db 951 CCTCCCTGTGACATGCTAAATAATGTGATTTTCTTTTAACTGACCTTCAACAGTTACT 1010
Qy 261 AspYsIleIleIleMetYsLeuGIYleuGIYleuGIYIleGIYleuGIYGIInIleLeuSer 280
Db 1011 GATAAATAATAAATGAAAAAGAAAAAGAAAGAAAGATTTGAGAAAGGGCAATCCTGTCT 1070
Qy 281 SerAlaIaApleIleAlaIleThrValGIInHieThrMetAlaCysHieLeuValIYsArg 300
Db 1071 TCAGCAGCAGCATTTGCTGCCACAGTACACACAAATGCAATGCTTTGTGAAAAGA 1130
Qy 301 ThrHieArgAlaIleLeuPheCysYsGIInArgApleuLeuProGIInAsnAsnAlaVal 320
Db 1131 ACACATGGGGCTATTTCTGTTTGTAAAGCAGAGACTTGTACTTCAAAATATATCAGTA 1190
Qy 321 LeuValAlaSerGIYGIYValAlaSerAsnPheYIleArgArgAlaLeuGIInIleLeu 340
Db 1191 CTGTTTCATCTGTGTGTGTGTGCAAGTAATCTTATCCGACAGAGCTCTGAAAAATTTTA 1250
Qy 341 ThrAsnAlaThrgInCysThrLeuLeuCysProProArgLeuCysThraAsnArgIY 360
Db 1251 ACNAAAGCAACAGTGCATTTGTGTGTCTCTCCACAGCTATGCACTATATATGCTC 1310
Qy 361 IleMetIleAlaATPAAsnGIYIleGIuArgLeuArgAlaGIYleuGIYIleLeuHieAsp 380
Db 1311 ATTATGATGTGATGAATGTATATGAAGAATACCTGCTGTGGCATTTTACTATGAC 1370
Qy 381 IleGIuGIYIleArgYIYGIuProIYsCysProLeuGIYValAspIleSerYsGIYVal 400
Db 1371 ATAAAGGCAATCCGCTATCAACCAAAATGTCTCTTGGAGTAAATATCAAAAGAGTT 1430
Qy 401 GIYGIuAlaSerIleYsValProGIInleuYsMetGIuIle 414

Db 1431 GGAGAGCTTCCATTAAGTACCAATTAATAAGAGATA 1472
RESULT 3
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-651-722-1
Alignment Scores:
Pred. No.: 2,416-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-649-273-2 (1-414) x US-10-651-722-1 (1-2197)
Qy 1 MetLeuIleLeuThrIleYsThraIaGIYValPhePheYsProSerYsArgYsValIYr 20
Db 231 ATGCTAATCTTGAATCACTGACAGAGATTTTAAAAACATCAAAAGAAAGTTAT 290
Qy 21 GIuPheLeuArgSerPheAsnPheHieSProGIYThrLeuPheLeuHieSlyIleValLeu 40
Db 291 GAATTTTAAAGAGTTTAAATTTTCACTCTGAAACATTTCTTCAATAAATAGATTTG 350
Qy 41 GIYIleGIuThrSerCysAplePheThraIaIaIaValaIaApleGIuThrgIYAsnVal 60
Db 351 GGAATTTGAAACTAGTGTGATGATACAGCAGCTGTGTGTGAGTGAACCTGAAAAATGTG 410
Qy 61 LeuGIYGIuAlaIleHieSerGIInThrgIuValHieLeuYsThrGIYIleValPro 80
Db 411 TTGGAGAGAAAGCAATACATTCCTCAAACTGAAGTTCAATTAAAAACAGTGGATGTCTCT 470
Qy 81 ProAlaIaGIInGInLeuHieSargGIuAenIleGIInargIleValGIInGIuAlaLeuSer 100
Db 471 CCAGCAGCTCAACAGCTTCACAGAGAAAATATTCAACGAATGATGACAGAACTCTTTCT 530
Qy 101 AlaseerGIYValSerProSerApleuSerAlaIleAlaThrIleYsProGIYleu 120
Db 531 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTGCACTACCAATAAACAGAGACTT 590
Qy 121 AlaleuSerLeuGIYValGIYleuSerPheSerLeuGIInleuValGIYInleuYsYs 140
Db 591 GCTTTAAGCTGGAGGTGGCTTATCATTTTACCTTACAGCTGTGAGACAGTTAAAG 650
Qy 141 ProPheIleProIleHieHieSmetGIuAlaHieAlaLeuThrIleArgLeuThraenYs 160
Db 651 CCATTTCATTCCTCATTCATCATATGAGGCTCATGCACTTACTATTAGTTGACCAATATA 710
Qy 161 ValGIuPheProPheLeuValleuLeuIleSerGIYGIYHieCysleuLeuAlaleuVal 180

Db	711	GGAGATTTCCTTTTAACTTTTAACTTTCTGGAGCTACGTCTGTGGCATTACTT	770
Qy	181	GINGLIYALSERAPPELEULEULEUGLYLSERLEUASPILEALIPROGLYASPHET	200
Db	771	CAAGGAGTTTCAGATTTTCTCTCTTGTGGAAGTCTTTGGACATGTGACCAAGTGAATG	830
Qy	201	LEUASPIYSLALAAATGATGLEUSERLEULEULEYSHISPROGLUCYSSERTHETSER	220
Db	831	CTTAAACAAGGTGCAAGAAAGACTTTCTTTTAAATAAACATCCAGATGCTCCACCAATGAT	890
Qy	221	GLYGLYLSALALEGLUHLIELEUALALYSGINGLYASATGPHESIPHEASPIILEYLS	240
Db	891	GGTGGAAAGCCATGAAACATTTGGCCAAACAGAAATAGATTTCATTTTGACATCAAA	950
Qy	241	PROPROLEUHLISHIALALYASNCYASAPPHESERTHERGILEUGLNHLISVALIHR	260
Db	951	CTCTCCCTTGCACTCATGTCAAAAATTTGATTTTTCTTTTAACTGACCTTCAACACCTTACT	1010
Qy	261	ASPIYSLILEIEMETLYSLGULYSGULYGLUGLYLIEGLULVSGLYGLNILELEUSER	280
Db	1011	GATPAAATPAAATPAAAGAAAAGAAAAGAAAGATTTGAGAAAGGAGGAGCAATCTGTCT	1070
Qy	281	SERIALAASPILEALALALATHRVALGLNHLISTHRMECALCYSHISLEUVALYSAATG	300
Db	1071	TCACGACGACGACATTTGCTGCCACAGTACAGCACCAATGGCATCTGTGAAAGA	1130
Qy	301	THRHSATGALALEULEUPHESYSLYSGULNATGAPLEULEUPROGLNASHNALVALI	320
Db	1131	ACACATCGGGCTATCTGTCTTTTGTAAAGCAGAGAGACTTGTAACTCAAAATPAAATGCACTA	1190
Qy	321	LEUVALIASERGLYGLYVALIASERASNPHERYILEATGATGALALEUGLULEULEU	340
Db	1191	CTGTTTGCACTTGTGTGTGTGTGCAATGATCTTATATCCGACAGACTCTGGAAATTTTA	1250
Qy	341	THRASHALATHRGINCYSYTHREULEUCYSPROPROATRGLEUCYSYTHRASPSNGLY	360
Db	1251	ACAAACGCAACACAGTGCACTTTGTTGTCTCTCTCCCAAGCTATGACATGATATATGCG	1310
Qy	361	ILEMETILEALATPASHNGLYLIEGLUATGLEUATGALAGLYLEUGLYYILEUHLISAP	380
Db	1311	ATTATGATTTGATGAAATGATTAAGAAAGACTACCTGTGGCTTGGGCAATTTTACATGAC	1370
Qy	381	ILEGLUGLYLEATGTYRGULUPROLYSCYSPROLEUGLYVALASPIILESERIYSGLYVAL	400
Db	1371	ATGAAAGCATCCGCTATGAAACCAAAATGTCTCTTGGAGTGAATATCAAAAGAAATT	1430
Qy	401	GLYGLUALASERILEYSLYVALIPROGLNDEUYSMETGULILE	414
Db	1431	GGAGAGCTTCCATPAAAGTACACCAATPAAAAATGGAGATA	1472
RESULT 4			
US-10-067-443-21			
Sequence 21, Application US/10067443			
Publication No. US20030082782A1			
GENERAL INFORMATION:			
APPLICANT: Bristol-Myers Squibb Company			
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED			
FILE REFERENCE: D0073 NP			
CURRENT APPLICATION NUMBER: US/10/067,443			
PRIORITY FILING DATE: 2002-02-05			
PRIORITY FILING DATE: 2001-02-05			
PRIORITY APPLICATION NUMBER: US 60/266,518			
PRIORITY FILING DATE: 2001-04-10			
NUMBER OF SEQ ID NOS: 71			
SOFTWARE: PatentIn version 3.0			
SEQ ID NO 21			
LENGTH: 1387			
TYPE: DNA			
ORGANISM: homo sapiens			
US-10-067-443-21			

Alignment Scores:	1,84e-243	Length:	1387
Pred. No.:	2090.50	Matches:	412
Score:	93.85%	Conservative:	0
Percent Similarity:	93.85%	Mismatches:	2
Best Local Similarity:	98.38%	Indels:	25
Query Match:	14	Gaps:	1
DB:	14		
US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)			
QY	1 MetLeuIleLeuThrThyThraIaGlyValI	PhePheLysProSerLysaRgIysValI	Tyr 20
Db	24 ATGCTAAATCTTGAACATCAAGAGCTGCAAGAGAGTTT	TTTAAACATCAAAAGAGAGTTTAT	83
QY	21 GIuPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu		40
Db	84 GAATTTTAAAGAAAGTTTATATTTTATCTCTGAACACATATTTCTTATATAAATAGATTC		143
QY	41 GIYIIeGIuThrSerCysAspAspThrAlaAlaValValAspGIuThrGlyAsnVal		60
Db	144 GGAATVTGAACACTAGTGTGATGATACAGAGCTGCTGTGTGTGTGAATGAACCTGAATATGTG		203
QY	61 LeuGIYGIuaIaIaHisSerGIuThrGlyValHisLysLysThrGlyGlyIleValPro		80
Db	204 TTGGAGAAAGCAATACATTTCCAAACTGAAGTTCACTTTAAAAACAGGTGGATGTTCT		263
QY	81 ProAlaAlaGInGInLeuHisArgGIuAsnIleGInArgIleValGInGInuaIaLeuSer		100
Db	264 CCAGCAGGCTCAACACCTTCAAGAGAAATATTCAGAAAGATGTACAGAGCCTTTTCT		323
QY	101 AlaSerGIYValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGIYLeu		120
Db	324 GCGAGTGAAGTCTCTCAAGTGAACCTTCAGAAATYGCACATCACTAAACCAAGACCTT		383
QY	121 AlaLeuSerLeuGIYValGIYLeuSerPheSerLeuGInLeuValGIYGIuLeuLysLys		140
Db	384 GCTTTAAAGCTTGGAGGTGGCTTATCATATTACTTTCACCTGGTAGACGTTAAAAAG		443
QY	141 ProPheIleProIleHisIleMetGIuAlaHisAlaLeuThrIleArgLeuThrAsnLys		160
Db	444 CCATTTCATTTCCATTATCATATATGAGAGCTCATGACCTACTATTAAGTTGACCAATAAA		503
QY	161 ValGIuPheProPheLeuValLeuLeuIleSerGIYGIYHisCysLeuLeuAlaLeuVal		180
Db	504 GTAGAGATTTCTTTTATTAGTTCTTTTGAATTTCTGGAGGTCACGTGTGTTGGCATTAAGTT		563
QY	181 GInGIYValIleSerAspPheLeuLeuLeuGIYLysSerLeuAspIleAlaIleProGIYAspMet		200
Db	564 CAAGAGTTTCAGATTTTCTGCTTCTGGAAAGTCTTGGACATACACACAGGTGACATG		623
QY	201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGIuLysSerThrMetSer		220
Db	624 CTTGACAAAGTGGCAAGAAAGCTTTCTTTAAATAAACATCCAGAGTGCTCCACATGAGT		683
QY	221 GIYGIYLysAlaIleGIuHisLysLeuAlaLysGInGIYAsnArgPheHisPheAspIleLys		240
Db	684 GGTGGGAAAGCCATAGAGCATTTGGCCAAACAGAAATAGATTTCATTTTGCATCAAAA		743
QY	241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGIYLeuGInHisValThr		260
Db	744 CTTCCCTTGCATCAATGCTAAATATGTGATTTTCTTTTAACTGGACTTCAACACGTTACT		803
QY	261 AspLysIleIleMetLysLysGIuLysGIuLysGIYIle-----		273
Db	804 GATTAATAATATATGAAAGAAAGAAAGAGAGATATTTCTTAATTAGTAAAGTTGAA		863
QY	274 -----		GIuLys 275
Db	864 CAGATTAATATTCCTGGATTGTCCTAAATAAGCTCATTTCTGACAGGTATGAGAG		923
QY	276 GIYGIuIleLeuSerSerAlaAlaAspIleAlaAlaThrValGInHisThrMetAlaCys		295
Db	924 GGGCAAACTCTCTTTCAGACAGACATTTGTGTCACAGTATCAAGACACAAATGGCATGT		983

QY 296 HisleuValIySargThrHisArgAlaIleuPheCylsGlnArgAspLeuPro 315
DB 984 CATCTTGAAAAGAACATCGGGCTATTCGTTTGTGAAGAGAGACATTTGTAACCT 1043
QY 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTrpIleArgArg 335
DB 1044 CAAATAATATGACGTACTGGTTCATCTGGTGTGTGCGAAGTAACCTTATATCCGACA 1103
QY 336 AlaIeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProAspGlu 355
DB 1104 GCTCTGGAATTTTAAACAAACGACACAGTCACCTTGTGTGTCTCTCCCAACACTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375
DB 1164 TGCACGTAAATGCGATTATGATTCATGCAATGATGATTAAGAACTACGTGTGCTTG 1223
QY 376 GlyIleLeuHisAspIleGluGlyIleArgTrpGluProLysCysProLeuGlyValAsp 395
DB 1224 GGCATTTTACATGACATGAAAGGACATCCGCTATGAACCAAAATGTCTCTTGGAGTAGAC 1283
QY 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1284 ATATCAAAAGAGTTGAGAGGCTTCATMAAAGTACCAATMAATGAGATG 1340

RESULT 5
US-10-649-273-21
Sequence 21, Application US/10649273
Publication No. US20040043407A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
FILE REFERENCE: D0073 CNT
CURRENT APPLICATION NUMBER: US/10/649,273
CURRENT FILING DATE: 2003-08-27
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 21
LENGTH: 1387
TYPE: DNA
ORGANISM: homo sapiens
US-10-649-273-21

Alignment Scores:
Pred. No.: 1,84e-243 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: 17 Gaps: 1

US-10-649-273-2 (1-414) x US-10-649-273-21 (1-1387)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTrp 20
DB 24 ATGCTAAATCTTACTTAAGACTGACAGAGTTTATTTTAAACCAATCAAAAGGAAGCTTAT 83
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValIleu 40
DB 84 GAATTTTAAAGAACTTTTAAATTTTCACTCGAAACACATATTTCTTCAATAAATAGTATTG 143
QY 41 GlyIleLeuHisAspIleGluGlyIleArgTrpGluProLysCysProLeuGlyValAsp 395
DB 144 GGAATTCAAATCTTACTTAAGAGCTTCAAGCTCTGTGTGCAATGTAAGCTGGAATGTG 203
QY 61 IeSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1284 ATATCAAAAGAGTTGAGAGGCTTCATMAAAGTACCAATMAATGAGATG 1340

DB 204 TTGGAGAAAGCAATACATTCACCAACTGAAGTTCATTTAAAAAACAGGTGGATGTTCT 263
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 264 CACAGAGCTCAACAGCTTCAACAGAAATAATTCAGAAATAGTACAAAGAGCTTTTCT 323
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 324 GCGAGTGAAGTCTCTCAACATGACCTTCAGCAATTCGCACTACCAATAAACAGAGCTT 383
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleLeuLys 140
DB 384 GCTTTAAGCTCGGAGAGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAAAG 443
QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 444 CCATTAATTCCTCATCATCATATGAGGCTCATGACCTTATGATGTTAGCAACAAATAA 503
QY 161 ValGluPheProPheLeuValIleLeuLeuIleSerGlyHisCysLeuLeuAlaLeuVal 180
DB 504 GTRGAATTCCTTTTATGTTAGTTCTTTGATTTCTGGAGTCACTGTCTTTGGCATTAAGT 563
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 564 CAAGGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGACATAGCACAGGAGCATG 623
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
DB 624 CTTGACAAAGGTGCAAGAAAGACTTCTTAATTAACATCCAAAGTCTCCACACATAGT 683
QY 221 GlyIleLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 684 GGTGGAAAGCCATAGACATTTGGCCAAACAAAGAAATGATTTCAATTTGACATCAAA 743
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 744 CCGCCCTTCATCATGCTAAATTTGATTTCTTTTATCTGAGCTTCAACAGCTTACT 803
QY 261 AspLysIleIleMetLysLysGlyLysGluGlyIle----- 273
DB 804 GATAAATATATATGAAAAAGAAAAAGAAAGATATATTCTAAATAGTAAAGTTGAA 863
QY 274 -----GluLys 275
DB 864 CAGATTAATATTCCTGAGATTGCTCTTAAAAATAGCTGCTCATTTCTGACAGTATGAGAA 923
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295
DB 924 GGGCAATCTCTGCTTCAAGCAGACATTTGCTGCCACAGTACGACACAAATGCGATGT 983
QY 296 HisleuValIySargThrHisArgAlaIleuPheCylsGlnArgAspLeuPro 315
DB 984 CATCTTGAAAAGAACATCGGGCTATTCGTTTGTGAAGAGAGACATTTGTAACCT 1043
QY 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTrpIleArgArg 335
DB 1044 CAAATAATGACGTACTGGTTCATCTGGTGTGTGCGAAGTAACCTTATATCCGACA 1103
QY 336 AlaIeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProAspGlu 355
DB 1104 GCTCTGGAATTTTAAACAAACGACACAGTCACCTTGTGTGTCTCTCCCAACACTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375
DB 1164 TGCACGTAAATGCGATTATGATTCATGCAATGATGATTAAGAACTACGTGTGCTTG 1223
QY 376 GlyIleLeuHisAspIleGluGlyIleArgTrpGluProLysCysProLeuGlyValAsp 395
DB 1224 GGCATTTTACATGACATGAAAGGACATCCGCTATGAACCAAAATGTCTCTTGGAGTAGAC 1283
QY 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1284 ATATCAAAAGAGTTGAGAGGCTTCATMAAAGTACCAATMAATGAGATG 1340

```
RESULT 6
US-10-651-722-21
; Sequence 21, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEINASE, ME-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-21

Alignment Scores:
Pred. No.: 1,84e-243 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: 17 Gaps: 1

US-10-649-273-2 (1-414) x US-10-651-722-21 (1-1387)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db 24 ATGCAATCTTCACTGACAGAGCTGAGAGTTTATTTAAACATCAAAAGAAAGTTTAT 83
QY 21 GluPheLysArgSerPheAsnDheHisProGlyThrLysPheLysHisLysIleValLeu 40
Db 84 GAATTTTAAAGAAAGTTTATTTATTTTCACTGAAACACATATTTCTTCATATAAATAGATTTG 143
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValValAspGluThrGlyAsnVal 60
Db 144 GGAATTGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 203
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
Db 204 TTGGAGAGAGCAATACATTTCCAAACTGAACTTCATTTTAAACAGAGTGGATTTCTCT 263
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnIleValLeuSer 100
Db 264 CCAGAGCTCAACACACTTCACAGAGAAATATTTCAACGAATGTACAGAGAGCTTTCTCT 323
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 324 GCCAGTGAAGTCTCTCCCAAGTACCTTCAGCAATTTGCAACTACCAATTAACCCAGACTT 383
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 384 GCTTTAAGCTGGAGAGTGGGCTTTACATTTAGCTTACAGCTGTAGAGACGTTTAAAGAG 443
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 444 CCATTCATTTCCATTCATCATATGAGGCTCAATGACCTTACTATTAGGTTGACCAATATA 503
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyValHisCysLeuLeuAlaLeuVal 180
Db 504 GTAGAAATTCCTTTTATTTAGTTCTTTGATTTCTTGAGAGTCACTGTCTGTGGCATTAAGT 563
QY 181 GlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspIleAlaProGlyAspMet 200
|||||
```

```
Db 564 CAAGAGTTTCAGATTTTCTGCTTCTTGGAAGCTTTTGACATAGACAGGTCAGC 623
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db 624 CTTGACAGAGTGCAGAGAGACCTTCTTTATATTAACATCCAGAGTCTTCACATAGAT 683
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 684 GGTGGAGAAAGCCATAGACATTTGGCCAAACAGAAATAGATTTCAATTTGACATCAA 743
QY 241 ProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 744 CCTCCCTTCATCATCTCAAAATTTGATTTTCTTTACTGACCTTCAACAGCTTACT 803
QY 261 AspLysIleIleMetLysLysGlyLysGluGlnGlyIle----- 273
Db 804 GATTAATAATATATGAAAAAGAAAAAGAAAGATATATTTCTTAATAGTAAAGTTGAA 863
QY 274 -----GluLys 275
Db 864 CAGATAAATATTCCTGAGATTTGCTTAAATAGCTGCTCATTTCTGACAGTATGAGAG 923
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 295
Db 924 GGGCAATCCTGCTTCAAGCAGACATTTGCTGCAAGTACAGACACAAATGGCATGT 983
QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315
Db 984 CATCTTGAAAAAGAACACATCGGCTATTTCTTTTGTGTAAGACAGAGACTGTATCCT 1043
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAspAsnPheThrIleArgArg 335
Db 1044 CAAATATATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1103
QY 336 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355
Db 1104 GCTCTGAAATTTTAAACAAACGACACAGTGCACCTTGTGTGTCTCTCCCAAGACTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuAlaGlyLeu 375
Db 1164 TGCACGTATATATGCACTTATGATTTGATGATGATGATGATGATGATGATGATGATG 1223
QY 376 GlyIleLeuHisAspIleGluGlyIleArgGlyIleArgGlyIleProLysCysProLeuGlyValAsp 395
Db 1224 GGCATTTTACATACATAGAGGCAATCCGCTTATGAAACCAAAATGTCTCTTGAGTAC 1283
QY 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1284 ATATCAAAAGAAATTTGAGAGAGCTTCCATTAATAAGTACCAATTTAATAATGAGATA 1340

RESULT 7
US-10-012-140-6
; Sequence 6, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leibny, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
```


LENGTH: 1245
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-012-140-6

Alignment Scores:

Pred. No.: 3,13e-243 Length: 1245
 Score: 2088.00 Matches: 407
 Percent Similarity: 99.03% Conservative: 3
 Best Local Similarity: 98.31% Mismatches: 4
 Query Match: 98.26% Indels: 0
 DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-012-140-6 (1-1245)

QY 1 MetLeuIleLeuThrThrAlaGlyValPhePheIysProSerIlyValArgIleValTyr 20
 DB 1 ATGCTAATCTTGACTAAGACGAGAGAGTTTATTTTAAACATCAAAAAGAAAGTTTAT 60
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValIleu 40
 DB 61 GAATTTTAAAGAGTTTATTTTATCTCTGGAACACTATTTCTTCATTAATAGTATTG 120
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyValAsnVal 60
 DB 121 GGAATTGAACTAGTGTGATGATACACAGAGCTGCTGTGTGATGGAATGGAATGTG 180
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuIysThrGlyGlyIleValPro 80
 DB 181 TTGGAGAGAGCAATACATTTCCCAACTGAAGTTCAATTTAAAAACAGTGGATGTTCT 240
 QY 81 ProIlaIaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
 DB 241 CCACAGAGCTCAACAGCTTCAAGAAAATATTCACAGAAATGTACAAAGAGCTCTTCT 300
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
 DB 301 GCCAGTGAAGCTCTCCCAAGTACCTCTCAGCAATTGCACTACCACTAAACAGAGCTT 360
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIysIys 140
 DB 361 GCCTTAAGCCCTGGAGTGGCTTATCATTTAGCTTACAGCTGTGTGAGCAAGTTAAAG 420
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160
 DB 421 CCATTCAATCCCATTCATCATATGAGGCTCATGACTATCATTTAGGTGACCAATAAA 480
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 DB 481 GTAGAAATTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 540
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerIleAspIleAlaProGlyAspMet 200
 DB 541 CAAGGAGATTTCAGATTTTCTGCTCTTGGAAAGCTTTTGGACATAGCACAGTACATG 600
 QY 201 LeuAspIysValAlaAlaGlyArgLeuSerIleIleValHisProGluCysSerThrMetSer 220
 DB 601 CTGACCAAGGTGGCAGAGAGCTTTCTTAAATAAACATCCAGAGTGTCCCACTGAGT 660
 QY 221 GlyGlyValValAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240
 DB 661 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAGAAATAGATTTTCATTTTGAACATAA 720
 QY 241 ProProLeuHisHisAlaAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 DB 721 CTTCCCTTGATCATCTGTAATAATTTGATTTTCTTTTATCTGCACTTCAACACCTTACT 780
 QY 261 AspIysIleIleMetIysLeuGluIysGlyGlyGlyIleGlyIysGlyGlnIleLeuSer 280
 DB 781 GATTAATAAT 840
 QY 281 SerIlaIaAspIleAlaIaIaThrValGlnHisThrMetAlaCysHisLeuValIlyArg 300

DB 841 TCAGCAGACAGACATTCCTGCCACAGTACAGACACACATGCAATGCTATCTTGTAAGAA 900
 QY 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnIlaAsnAlaVal 320
 DB 901 ACACATCGGCTATCTGTTTGTATGACAGAGAGACTGTGTACTCAAAATTAATGACGTA 960
 QY 321 LeuValAlaSerGlyValAlaIleAsnPheThrIleArgArgAlaLeuGluIleLeu 340
 DB 961 CTGGTTCATCTGCTGTGTGTGCGCAATTACTTATATCGCAGAGAGCTTGGAAATTTTA 1020
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
 DB 1021 ACAACGCAACACAGAGCACTTGTGTGTCTCTCCAGACTATGCACTGATTAATAGGC 1080
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
 DB 1081 ATTATATATGCAATGAGATGTATTAAGACATAGCTGCTGGCATTTTAAACATGAC 1140
 QY 381 IleGluGlyIleArgGlyGluProIysCysProLeuGlyValAspIleSerIysGluVal 400
 DB 1141 ATGAGAGCATCGCTATACCAAAATGTCTCTTGAGATACATATCAAAAGAGTT 1200
 QY 401 GlyGluAlaSerIleIysValProGlnLeuIysMetGluIle 414
 DB 1201 GGAGAGCTTCCATTAAGATACCAATTAATAATGAGATA 1242

RESULT 8
 US-10-012-140-4
 ; Sequence 4, Application US/10012140
 ; Publication No. US20030009017A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Leiby, Kevin R.
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; APPLICANT: Glucksmann, Marcia A.
 ; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
 ; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
 ; FILE REFERENCE: 381552004900
 ; CURRENT APPLICATION NUMBER: US/10/012,140
 ; PRIOR FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: 60/246,768
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: 60/246,772
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: 60/249,185
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 1820
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (146)...(1390)
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1820)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-10-012-140-4

Alignment Scores:
 Pred. No.: 5.69e-243 Length: 1820
 Score: 2088.00 Matches: 407
 Percent Similarity: 99.03% Conservative: 3
 Best Local Similarity: 98.31% Mismatches: 4
 Query Match: 98.26% Indels: 0
 DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-012-140-4 (1-1820)

QY 1 MetLeuIleLeuThrThrAlaGlyValPhePheIysProSerIlyValArgIleValTyr 20

QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
 DB 584 CCAGAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATGTCACAAAGACTCTTTCT 643
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 DB 644 GCCAGTGAAGTCTCTCCAAAGTGAACCTTCAGCAATGCAACTACCAATAAACAGAGACTT 703
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
 DB 704 GCTTAAAGCTGGAGGAGGCTTACATTACATTACAGCTGGTAGAGAGATTAAAAAG 763
 QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 DB 764 CCATTCAATCCCATTCATCATATGAGGCTCATGCACTTACTATTAAGTTGACCAATAAA 823
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 DB 824 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACGTCTGTGGCAATTAGTT 883
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 DB 884 CAAGAGATTCAAGATTTCTGCTTGTGAAAGCTTTGACATAGCACAGGTGACATG 943
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
 DB 944 CTTTACAAAGGTGGCAAGAAAGCTTCTTTAATTAACATCCAGAGGCTCCACAAAGACT 1003
 QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLys 240
 DB 1004 GGTGGAAAGCCATAGAACTTTGGCCAAACAAAGAAATAGATTTCATTTTGAATCAAA 1063
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 DB 1064 CCTCCCTTGCAATGCTTAAATAATGTGATTTCTTTTAACTGCACTTCAACACCTTACT 1123
 QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLysGlyGlnIleLeuSer 280
 DB 1124 GATTAATAATTAATGAAGAAAGAAAGAAAGAAAGATTTGAAGAGGGGCAAAATCCCTGCT 1183
 QY 281 SerAlaAlaSerIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
 DB 1184 TCACAGACAGACATTTGCTGCCAGTACAGACACAAATGCAATGCTTGTGAAAACA 1243
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
 DB 1244 ACAATCGGGCTATTTCTGTTGTATGACAGAGACTTGTATCTCAAAATATATCAGTA 1303
 QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
 DB 1304 CTGGTTCATCTGGTGGTGGTGGCAAGTAACTTGTATCCGCAAGCTCTGGAAAATTTTAA 1363
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGly 360
 DB 1364 ACAAAACCAACAGTGCATTTGTTGTCTCTCCCAAGCTATGCACTGATATATGCGC 1423
 QY 361 IleMetIleAlaTyrAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
 DB 1424 ATATATGATTCGA----- 1435
 QY 381 IleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlnVal 400
 DB 1436 -----TGATGTCTCTTGGAGTAGAATATCAAAAGAGTT 1471
 QY 401 GlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 414
 DB 1472 GGAGAACTTCATTAAGTACCAACAATTAATAAATGAGATA 1513
 RESULT 10
 US-10-723-860-7447
 ; Sequence 7447, Application US/10723860
 ; Publication No. US20040253606A1

; GENERAL INFORMATION:
 ; APPLICANT: Aziz, Nacasha
 ; APPLICANT: Ginsburg, Wendy M.
 ; APPLICANT: Zlotnick, Albert
 ; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
 ; FILE REFERENCE: 05882.0193.NPUS01
 ; CURRENT APPLICATION NUMBER: US/10/723,860
 ; PRIOR FILING DATE: 2003-11-26
 ; PRIOR APPLICATION NUMBER: 60/429,739
 ; PRIOR FILING DATE: 2002-11-26
 ; NUMBER OF SEQ ID NOS: 8393
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 7447
 ; LENGTH: 2890
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (646)..(657)
 ; OTHER INFORMATION: n is a, c, g, or t
 ; US-10-723-860-7447

Alignment Scores:
 Pred. No.: 3,93e-225 Length: 2890
 Score: 1944.00 Matches: 386
 Percent Similarity: 93.24% Conservative: 0
 Best Local Similarity: 93.24% Mismatches: 4
 Query Match: 91.48% Indels: 24
 DB: Gaps: 1
 US-10-649-273-2 (1-414) x US-10-723-860-7447 (1-2890)

QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
 DB 1001 ATGCTAATCTTGAACATACTGACAGAGATTTTAAACATCAAAAGAAAGATTAT 1060
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
 DB 1061 GAATTTTAAAGATTTTAAATTTTCAATCCGGAACACTATTTCTCAATAAATATGATTTG 1120
 QY 41 GlyIleGlnThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60
 DB 1121 GGAATTTGAACCTAGTTGTATGATATACAGAGCTGCTGTGTGATGAACTGGAATATGTG 1180
 QY 61 LeuGlyGlnAlaIleHisSerGlnThrGlnValHisLeuLysThrGlyIleValPro 80
 DB 1181 TTGGAGGAAGCAATACATTTCCAAATCGAAGTTCATTTAATAAACAGGTGGATTTCTCT 1240
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
 DB 1241 CCAGAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATATGACAAAGCTCTTCT 1300
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 DB 1301 GCCAGTGAAGTCTTCCCAAGTACCTTCAGCAATGCAACTACATTAAGTTGATTAATAA 1360
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
 DB 1361 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTCAAGCTGTATGACAGCTTAAAAAG 1420
 QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 DB 1421 CCATTCAATCCCATTCATCATATGAGGCTCATGCACTTACTATTAAGTTGATTAATAA 1480
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 DB 1481 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACGTCTGTGGCAATTAGTT 1540
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 DB 1541 CAAGAGTTTCAGATTTTCTGCTTGTGAAAGTCTTTGACATAGCACAGGTGACATG 1600

```

Qy 201 LeuAspIysValAlaArgArgLeuSerLeuIleIleYsHisProGluCysSerThrMetSer 220
Db 1601 CTTCACAAAGTGGGAGAAAGACTTCCTTAATATAACATCCAGATGCTCCACCACTGAAT 1660
Qy 221 GlyGlyIysValAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHisPheAspIleYs 240
Db 1661 GGTGGGAAAGCCATAGAAACATTTGGCCAAACAAAGGAAATAGATTTCATTTGACATCAAA 1720
Qy 241 ProProLeuHisHisAlaIleYsAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 1721 CCTCCCTTGACATCAAGCTTAAATAATTTGATTTTCTTTTACCTGACCTTCAACACGTTACT 1780
Qy 261 AspIysIleIleMetIleYsGlyIleYsGluIleGluGlyGlyGlnIleLeuSer 280
Db 1781 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1840
Qy 281 SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisIleValIysArg 300
Db 1841 TCACACAGACAGATGCTGCCACAGTACAGCACACAAATGGCATGTCATCTTGAAAAAG 1900
Qy 301 ThrHisArgAlaAlaIleLeuPheCysIleYsGlnArgAspIleLeuProGlnAsnAlaVal 320
Db 1901 ACACATCGGCTATTCCTTTTGTAAAGACAGAGACTTGTATCTCAAAATTAATGCAGTA 1960
Qy 321 LeuValAlaSerGlyGlyValAlaIleSerAsnPheTyrlleArgArgAlaIleuGluIleu 340
Db 1961 CTGTTGATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2020
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
Db 2021 ACAACGCAACACAGTGCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2080
Qy 361 IleMetIleAlaIlePheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHis 380
Db 2081 ATTAATGATTTGCA-----TGATGTCCTCTTGAGTAGACATATCAAAAGATT 2092
Qy 381 IleGluGlyIleArgTyrlleGluProIleCysProLeuGlyValAspIleSerIleGlyVal 400
Db 2093 -----TGATGTCCTCTTGAGTAGACATATCAAAAGATT 2128
Qy 401 GlyGluAlaSerIleYsValProGlnLeuIleYsMetGluIle 414
Db 2129 GGAGAGCTTCATTAAGTACCAATTAATAAAGGAGATA 2170

RESULT 11
US-10-120-988-177
; Sequence 177, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Kyle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyan
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radole T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/120,988
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
US-10-120-988-177

```

Alignment Scores:

```

Pred. No.: 1,199-201 Length: 1416
Score: 1747.00 Matches: 340
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.71% Mismatches: 0
Query Match: 82.21% Indels: 0
DB: 17 Gaps: 0

```

US-10-649-273-2 (1-414) x US-10-120-988-177 (1-1416)

```

Qy 74 IysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
Db 280 AGAACAGGGGGATTTGTTCTCCAGACGCTCAACAGCTTCACAGAGAAATATTAACGCA 339
Qy 94 IleValGlnGluAlaLeuSerAlaSerGlyValIleSerProSerAspIleSerAlaIle 113
Db 340 ATAGACAAAGAGCTTTCTGCGAGTGAAGTCTCCAGATGACCTTCAGCAAAATGCA 399
Qy 114 ThrThrIleYsProGlyLeuAlaLeuSerIleGlyValGlyLeuSerPheSerLeuGln 133
Db 400 ACTACCATTAACACAGACTTGTCTTAAGCTGGAGTGGCTTATCATTTACTTACCTTACG 459
Qy 134 LeuValGlyGlnLeuYsIleYsPheProPheIleProIleHisIleGluAlaHisAlaLeu 153
Db 460 CTGTAGACAGATTAATAAAGCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 519
Qy 154 ThrIleArgLeuThrAsnIysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173
Db 520 ACTAATTAAGTTACCAATTAAGTAAATTTCTTTTAACTTTTGAATTTGAGGT 579
Qy 174 HisCysLeuLeuAlaLeuValGlnGlyValIleSerAspPheLeuLeuGlyIysSerLeu 193
Db 580 CACTGCTCTTGGCATTTAGTTCAAGAGATTTCAGATTTCTGCTTCTTGAAAAGCTTTG 639
Qy 194 AspIleAlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleYsHis 213
Db 640 GACATAGCACCGAGTGCATGCTTGAACAAGTGGCAAGAACTTCTTAATTAATAACAT 699
Qy 214 ProGluCysSerThrMetSerGlyGlyIysAlaIleGluHisLeuAlaIysGlnGlyAsn 233
Db 700 CCAGAGTGTCCACCACTGAGTGGGAAAGCATTTGGCAAAACAAGAAAT 759
Qy 234 ArgPheHisPheAspIleYsProProLeuHisHisAlaIysAsnCysAspPheSerPhe 253
Db 760 AGATTTCAATTTGACATCAACCTCCCTTGACATGCTTAATAATTTGATTTTCTTTT 819
Qy 254 ThrGlyLeuGlnHisValThrAspIysIleIleMetIleYsGlnIleYsGlnGlyIle 273
Db 820 ACTGACCTTCAACAGCTTACTGATTAATAATTAATAAAGGAAAGGAGGATTT 879
Qy 274 GlyIysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMet 293
Db 880 GAGAAAGGGCAATCTCTGCTTCAGCAGCAGCATTTGCTCCACAGTACGACACATG 939
Qy 294 AlaCysHisIleValIleYsArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeu 313
Db 940 GATGTCATCTTGTAAGAAACACATCGGCTATTTCTGTTTGTAAAGAGAGACTTG 999
Qy 314 LeuProGlnAsnAlaValIleuValAlaSerGlyGlyValAlaIleSerAsnPheTyrlle 333
Db 1000 TTACCTCAAAATTAATCACTACGTTGCTGCGTGGTGGCAAGTAACTTCTATATC 1059
Qy 334 ArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
Db 1060 CCAGAGCTCTGGAATTTTAAACAACGACACATGCGCTTTGTTGTCCTCCCTCCC 1119
Qy 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaIlePheAsnGlyIleGluArgLeuArgAla 373
Db 1120 AGACTATGCACTGATTAATGCAATTAATGATGCAAGAAAGTAAATTAACATACGCT 1179
Qy 374 GlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrlleGluProIysCysProLeuGly 393
Db 1180 GGTGGGCAATTTTACATGACATAGAGGCAATCCGCTATGAAACCAAAATGTCCTTTGGA 1239

```

```

Oy      394  VALAPELLeserlysgIuValGIyGluAAserIlelysvaIProGInleuIysmetGlu 413
        |||||
Db      1240  GTAGACATATCAAAAGAAAGTTGGAGAAAGCTTCATATAAGTACCACAATTTAAAAATGAG 1299

Oy      414  Ile 414
        |||
Db      1300  ATA 1302

RESULT 12
: Sequence 23, Application US/10067443
: Publication No. US20030082782A1
: GENERAL INFORMATION:
: APPLICANT: Bristol-Myers Squibb Company
: TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
: TITLE OF INVENTION: SPINAL CORD, MP-1
: FILE REFERENCE: D0073 NP
: CURRENT APPLICATION NUMBER: US/10/067,443
: CURRENT FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US 60/266,518
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 60/282,814
: PRIOR FILING DATE: 2001-04-10
: NUMBER OF SEQ ID NOS: 71
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 23
: LENGTH: 1526
: TYPE: DNA
: ORGANISM: homo sapiens
: US-10-067-443-23

```

Alignment Scores:				
Pred. No.:	1.49e-157	Length:	1526	
Score:	1385.00	Matches:	267	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	65.18%	Indels:	0	
DB:	14	Gaps:	0	
US-10-649-273-2 (1-414) x US-10-067-443-23 (1-1526)				
QY	148 MetGIuAlaHisAlaLeuThrIleArgLeuThrAsnLysValGIuPheProPheLeuVal 167			
Db	1 ANGGAGGCTCATGCACTTACTATTAGGTGACCAATAAGTGAATTTCTTTTATGTT 60			
QY	168 LeuLeuIleSerGIyGIyHisCysLeuLeuAlaLeuValGIyValSerAspPheLeu 187			
Db	61 CTTTGATTTCTGGAGGCTCACTGCTCTGTGGCATTAGTTCAGAGGATTCAGATTTTCTG 120			
QY	188 LeuLeuGIyLysSerLeuAspIleAlaProGIyAspMetLeuAspLysValAlaArg 207			
Db	121 CTTTGTGAAAGCTTTTGACATAGCACAGGTGACATGCTTGACAAAGGTGCMAAAGA 180			
QY	208 LeuSerLeuIleLysHisProGIuCysSerThrmSetSerGIyLysAlaIleGIuHis 227			
Db	181 CTTTCTTTAATAAACAATCCAGAGTCTCCACCAATGATGGTGGGAAGCCATTAGACAT 240			
QY	228 LeuAlaLysGIyGIyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247			
Db	241 TTGGGCCMAACAGGAATAATAGATTTCATTTTGACATCAAAACCTCCCTTGATCATGCTAA 300			
QY	248 AsnCysAspPheSerPheThrGIyLeuGIuHisValThrAspLysIleIleMetLysLys 267			
Db	301 AATTGTGATTTTCTTTTACTGCACTTCACACAGCTTACTGATTAATAATATGAAAAAG 360			
QY	268 GIuLysGIuGIyLysIleGIuLysGIyLysIleLeuSerSerAlaAlaAspIleAlaAla 287			
Db	361 GAAAAAGAGAGAGTATTGAGAGGGGCCAAATCTGTCTTCAGACGACGACATCTGCC 420			
QY	288 ThrValGIuHisThrMetAlaCysHisLeuValLysAspThrHisArgAlaIleLeuPhe 307			
Db	421 ACAGTAGACACACAATGAGCATGTCATCTGTGAAAGAAACAACATCGGGCTATTCTGTT 480			

QY	308	CyelyrGGlnArAspleuleuProGlnaMaMaLaValleuVaMaAserylVGIyVal	3277
DB	481	TGTAAAGCAGAGAGACTTGTACTCCAATAATATGCAGTCTGGTGCATCTGGTGGTGC	540
QY	328	AlaseraenPhaTylleArgrgMaaleuGluileuThraMaLaThrguCySTR	347
DB	541	GCAAGTAATTTATATCCGACAGACTCTGGAATTTTAAACAACGCAACACAGTGCAC	600
QY	348	LeuleuCySPProProArgrleuCySTRaSPaNGlyIlemerIleLaTRPaNGly	367
DB	601	TGTGTGTGTCTCTCTCCAGACTATGCATGTGATTAAGCATTTATGATTCATGAAATGGT	660
QY	368	IleGluArgrleuArGlaIyIleuGlyIleuHISaSPileGluIyIleArgrYrglu	387
DB	661	ATTGAAAGACTAGGTGGCTGGCTTGGCATTTTACATGACATTAAGGACATTCGGCTATGAA	720
QY	388	ProLySArSPProleuGlyValaSPileSerlySgluValGlyGluAlaSerIleLyVal	407
DB	721	CCAAATATGCTCTTGAGATGACATATCAAAAGAAATTGGAGAACCTTCATATAAAGTA	780
QY	408	ProGlnleuLySerMetGluIle	414
DB	781	CCACAATTAAATATGAGATA	801

```

RESULT 13
US-10-649-273-23
/ Sequence 23, Application US/10649273
/ Publication No. US20040043407A1
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
/ FILE REFERENCE: D0073 CNT
/ CURRENT APPLICATION NUMBER: US/10/649,273
/ CURRENT FILING DATE: 2003-08-27
/ PRIOR APPLICATION NUMBER: US 60/266,518
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: US 10/067,443
/ PRIOR FILING DATE: 2002-02-05
/ PRIOR APPLICATION NUMBER: US 60/282,814
/ NUMBER OF SEQ ID NOS: 71
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 23
/ LENGTH: 1526
/ TYPE: DNA
/ ORGANISM: homo sapiens
US-10-649-273-23

Alignment Scores:
Pred. No.: 1,496-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-23 (1-1526)
QY 148 MetGluAaHsSalAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 167
Db 1 ATGGAGGCTCAGCAGCACTTACTATTGAGTTGACCATTAAGTAGAATTTCTTTTAACTT 60
QY 168 LeuLeuIleSerGlyGlyVHISCySLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db 61 CTTTGAATTTCTGGAGGCTCAGTCTCTGTGGCATATGATTCAGAGGATTCACATTTTCTG 120
QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
Db 121 CTTCTGAAAGAGCTCTTGACATACACACAGGTGACATGCTTGACACAGGTGGCAGAGACA 180
QY 208 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 227

```

Db 181 CTTCTTAAATPAAACATTCAGAGTCTCCACCATGATGTTGGAAAGCCATAGACAT 240
Qy 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisIleLys 247
Db 241 TTGGCCAAACAGAAATGATTTGATTTGACATCAACCTCCCTTGATCATGCTAA 300
Qy 248 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
Db 301 AATTGTGATTTTCTTTTACTGACTTCACACGTTACTGATTAATAATATGAAAAAG 360
Qy 268 GlnLysGlnGlnGlyIleGlnLysGlnGlnIleLeuSerSerAlaIleAspIleAla 287
Db 361 GAAAAAGAGAGAGATTTGAGAAAGGCGCAATCTGTCTTCAGACAGACATGCTGCC 420
Qy 288 ThrValGlnHisThrMetAlaCysHisLysValLysArgThrHisArgAlaIleLeuPhe 307
Db 421 ACAGTACGACACACATGCGATGTCATCTTGAAAAAGAACATCGGGCTATTCTGTT 480
Qy 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 327
Db 481 TGTAAAGCAGAGACTTGTGACTTCAAAATAATGCAATGCTGTCATCTGTGTC 540
Qy 328 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347
Db 541 GCAAGTAACTTCTATATCCGACAGAGCTCTGAAATTTTAAACCAACGACACAGTGCAT 600
Qy 348 LeuLeuCyProProProArgLeuCyThrAspAsnGlyIleMetIleAlaIleAsnGly 367
Db 601 TTGTTGTGTCCTCCCTCCACAGACTATGCAATGATTAATGATTAATGATTAATGAT 660
Qy 368 IleGlnArgLeuArgAlaGlyLeuGlnIleLeuHisAspIleGlnGlyIleArgTyrGln 387
Db 661 ATTGAAGACATACGCTGCTGCTGGCATTTTACATGACATAGAAAGCATCCGCTATGAA 720
Qy 388 ProLysCyProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 407
Db 721 CCAAAATGTCCTTGTGAGTACATATCAAAAGAAAGTTGAGAGAGCTTCATPAAAGTA 780
Qy 408 ProGlnLeuLysMetGlnIle 414
Db 781 CCACAAATPAAATGAGATA 801

RESULT 14.
US-10-651-722-23
Sequence 23, Application US/10651722
Publication No. US20040048302A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
FILE REFERENCE: D0073 DIV
CURRENT APPLICATION NUMBER: US/10/651,722
PRIOR FILING DATE: 2003-08-29
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 23
LENGTH: 1526
TYPE: DNA
ORGANISM: homo sapiens
US-10-651-722-23

Alignment Scores:
Pred. No.: 1,496-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2 (1-414) x US-10-651-722-23 (1-1526)
Qy 148 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 167
Db 1 ATGGAGGCTCAGTCACTTACTATTAGTTGACCAATTAAGTAAATTTCTTTTATGTT 60
Qy 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db 61 CTTTGTATTTCTGAGAGTCACTGTCGTGTCGCAATTGTTCAAGAGATTTTCAGATTTCG 120
Qy 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 207
Db 121 CTTCTTGAAGAGTCTTTTGACATAGCACAGGATGATCATGCTTGACAGAGGCGACAGAG 180
Qy 208 LeuSerLeuIleLysHisProGlnCySerThrMetSerGlyGlyLysAlaIleGlnHis 227
Db 181 CTTTCTTTAATPAAACATCCAGAGTGTCTCCACCATAGTGTGGAAAGCATPAGAACAT 240
Qy 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
Db 241 TTGGCCAAACAGAAATGATTTGATTTGACATCAACCTCCCTTGATCATGCTAA 300
Qy 248 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
Db 301 AATTGTGATTTTCTTTTACTGACTTCACACGTTACTGATTAATAATATGAAAAAG 360
Qy 268 GlnLysGlnGlnGlyIleGlnLysGlnGlnIleLeuSerSerAlaIleAspIleAla 287
Db 361 GAAAAAGAGAAAGTATGAGAAAGGCGCAATCTGTCTTCAGACAGACATGCTGCC 420
Qy 288 ThrValGlnHisThrMetAlaCysHisLysValLysArgThrHisArgAlaIleLeuPhe 307
Db 421 ACAGTACGACACCAATGCGATGTCATCTTGAAAAAGAACATCGGGCTATTCTGTT 480
Qy 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 327
Db 481 TGTAAAGCAGAGACTTGTGACTTCAAAATAATGCAATGCTGTCATCTGTGTC 540
Qy 328 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347
Db 541 GCAAGTAACTTCTATATCCGACAGAGCTCTGAAATTTTAAACCAACGACACAGTGCAT 600
Qy 348 LeuLeuCyProProProArgLeuCyThrAspAsnGlyIleMetIleAlaIleAsnGly 367
Db 601 TTGTTGTGTCCTCCCTCCACAGACTATGCAATGATTAATGATTAATGATTAATGAT 660
Qy 368 IleGlnArgLeuArgAlaGlyLeuGlnIleLeuHisAspIleGlnGlyIleArgTyrGln 387
Db 661 ATTGAAGACATACGCTGCTGCTGGCATTTTACATGACATAGAAAGCATCCGCTATGAA 720
Qy 388 ProLysCyProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 407
Db 721 CCAAAATGTCCTTGTGAGTACATATCAAAAGAAAGTTGAGAGAGCTTCATPAAAGTA 780
Qy 408 ProGlnLeuLysMetGlnIle 414
Db 781 CCACAAATPAAATGAGATA 801

RESULT 15
US-10-067-443-20
Sequence 20, Application US/10067443
Publication No. US20030082782A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 20
LENGTH: 14364
TYPE: DNA
ORGANISM: homo sapiens
us-10-067-443-20

Alignment Scores:

Pred. No.:	7,32e-132	Length:	14364
Score:	1186.50	Matches:	313
Percent Similarity:	32.30%	Conservative:	1
Best Local Similarity:	32.20%	Mismatches:	2
Query Match:	55.84%	Indels:	657
DB:	14	Gaps:	4

US-10-649-273-2 (1-414) x US-10-067-443-20 (1-14364)

QY	74	LySThrGlyGlyIleValProProAlaAgingInLeuHISArgGluAsnIleGlnArg	93
DB	10623	AGAACAGGTGGATTTCTTCCTCCAGCACTCAACAGCTTCAAGAGAAATTTCAACA	10682
QY	94	IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla	113
DB	10683	ATAGTACAAAGAGCTTTCTGCCAGTGAAGTCTCCAGAGTACCTCCAGCAATTGCA	10742
QY	114	ThrThrIleValProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln	133
DB	10743	ACTCCATTAACCAAGCACTGCTTTAAGCTGGAGTGGCTTATCATTTAGCTTACAG	10802
QY	134	LeuValGlyGlnLeuValLeuProPheIleProIleHisIleMetGluAlaHisAlaLeu	153
DB	10803	CTGGTAGGACGTTAAAGGCAATTCATCCCATTCATCATATGAGGCTCATGCACTT	10862
QY	154	ThrIleArgLeuThrAsnIleValGluPheProPheLeuValLeuLeuIleSerGlyGly	173
DB	10863	ACTATTAGGTGACCAATTAAGTAGAATTTCTTTTATAGTTCTTTGATTTCTGAGGT	10922
QY	174	HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeu	193
DB	10923	CACGTGCTGTTGGCATTTAGTTCAAGAGATTTCAGATTTTCTCTTCATGGAAGCTTTG	10982
QY	194	AspIleAlaProGlyAspMetLeuAspIleVal-----	204
DB	10983	GACATGACACCAAGGTGACATGCTTGACAAGGT-AATTAGAATTAAATTTCTCCATTTCTT	11041
QY	204	-----	204
DB	11042	TTTGTATGTTGTCCATTTCAACTRAGTAGCAATGATGTGCTACCACCATTCACCTTAA	11101
QY	204	-----	204
DB	11102	TATTTCTGAATTTTATCTTAGTAAACGAAAAAATTCACATATGCTGAGAAAAATAGA	11161
QY	204	-----	204
DB	11162	AAGAGTAGTACAAATTTATTAATCTTAGGCTTTCTTAATTAATGTAAGAGGTTCAT	11221
QY	204	-----	204
DB	11222	ATCTGTACATTAAGGCTGAATAATGTTGCAGATACGTTATGTATTTGCCAAATATGAT	11281
QY	204	-----	204
DB	11282	ATGTGAAGAAGACGTCTTGTAACTAACATATCTGCAAAAAAGTAAATAAGAGAAATAT	11341
QY	204	-----	204
DB	11342	ATATAGATTAAACATTAACCACTTAAAGATGCATACAGAGATTAATATCACAAATTAAT	11401
QY	204	-----	204

DB	11402	TACACCAACAGAGGTCCCCCCCCCACCCTTTGTTTATGATATACAGAGGCTTACT	11461
QY	204	-----	204
DB	11462	GCCATATATAGAAAAACAACAACAACAACAACAACAACACTGCTTCCACAGTGAATTA	11521
QY	204	-----	204
DB	11522	TAGAGTATAGGACAAAGTTCTTATTAATGACGTTATCATTAAGCAGTTATTTCTCACT	11581
QY	204	-----	204
DB	11582	TCAAGCCATTTTCCAACATAGAAAGACAAACATGACAGGGCAGTATGGCTCT	11641
QY	204	-----	204
DB	11642	TATTTGGGTGATCATTAAGAACAGGGTGTCTGCTTACCTGAATATCAGCTATAGTC	11701
QY	204	-----	204
DB	11702	TATATTGCCAAGTATAGCATGTTTATTCATTACAGGGGTTTTTGTGTTAGTAAT	11761
QY	204	-----	204
DB	11762	TTTCAATTAATTTCTTTCATCTTTTCGTTTCACAGATATTAATTTATAGACTTAAA	11821
QY	205	-----	204
DB	11822	AATATGTTCTTGTATAGTGGGACAGAGACCTTTCTTAATTAACAATCCAGAGTGTCC	11881
QY	218	ThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe	237
DB	11882	ACCATAGTGTGTGGAAAGCATATGACATTTGGCCAAACAGAAATATGATTCATTTT	11941
QY	238	AspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGln	257
DB	11942	GACATCAACCTCCCTGTCATCATGCTTAATAAATTTGATTTTCTTTACTGAGCTTCAA	12001
QY	258	HisValThrAspLysIleIleMetLysLysGlyGluGlu-----	271
DB	12002	CACGTACTGATTAATAATTAATGAAGAAAGAGAGAGATTAATTTCTAATAGT	12061
QY	272	-----	272
DB	12062	AAAGTTGAAACAGATAAATAATCTGTGATGTGCTTAAATAATGCTGCTCATTTCTGAGG	12121
QY	272	YIleGlyLysGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThr	292
DB	12122	TATTTGAGAGGGGCAATCTGTCTTCAGCAGACATTTGCTGCCAGTACAGCACAC	12181
QY	292	HisAlaCysHisIleValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAs	312
DB	12182	AATGGCATGTCATCTTGTGAAGAAACACATGGGGCTATTTCTTTTGTAAAGAGAGA	12241
QY	312	PleuLeuProGlnAsnAsnAlaValLeu-----	321
DB	12242	CTTGTACCTCAAAATTAATGACATCTGTGAAGTTTATCTCAATTTATGTAATAGTTA	12301
QY	321	-----	321
DB	12302	CACTTGCAATATGTTACTTTTTCACAAAGACCTTGACCTGTGTTAGATGAACAGAT	12361
QY	321	-----	321
DB	12362	CTTTATGCTTATGCTAGCCCTGACAGTATGAATTAATGACAGATAGGAAGACTAACG	12421
QY	321	-----	321
DB	12422	CCATTTCTGTACTAGTTGTAGCTTTATGGACAGCTGTATAGCTTATAGCACATA	12481
QY	321	-----	321
DB	12482	AGCTAATTTTGCATCTTCTGTGTGATTTAAAGAGGCTTACAAATAAGAAAGTAAT	12541

QY 321 ----- 321
Db 12542 GCAGTAAGCTGATCACTATTTTAGAAAAATAGGTGATTTCTTCATCTTGATGAA 12601
QY 321 ----- 321
Db 12602 ATCCCTTGTTGTTGTTTTTTTATAGCAAGTCAATTAGCAGTGGAGGTGAT 12661
QY 321 ----- 321
Db 12662 TCCACTTCGACACTAATGTGATTAAGTTCTGATATCCTATATGTACCAACC 12721
QY 321 ----- 321
Db 12722 AAAATCCCTTATATGCTTAAAGCCTTGACAAACATCTGTTAACTGATCTTAA 12781
QY 321 ----- 321
Db 12782 CTTTATTCATTTAAATAATTAATAAGTGGAAAAATGTTAAATGTAGTAATTCAT 12841
QY 321 ----- 321
Db 12842 AGATGGAATTTTACATGATATCAAGAATATTTTTCAGAGTTATGTATAAATGCA 12901
QY 321 ----- 321
Db 12902 CAAAAATAAAAAATTTGAGGCTTAAAAATGATGACTATGATGAATTAATTAATA 12961
QY 321 ----- 321
Db 12962 AATATTTAGATGAAGTTGGAGAAAAATATCAAAAAATGCTAGTAATGTTGTATGCTA 13021
QY 321 ----- 321
Db 13022 TTGAAATTATTAGTAATTTTCTTCCAAATTTTATTAATATAGATATGCTGACC 13081
QY 321 ----- 321
Db 13082 CATTAACCATCTCAAAATGGATGATTATTTATGTTAATGCTGATATTTTCTCCAGG 13141
QY 321 ----- 321
Db 13142 TTTAATTAGCAGCTTGCTCATATCCATATAGATATTTGTTTCTCAATTCCT 13201
QY 322 ---ValAserGlyValAlaSerAsnPhetYrIleArgAlaLeuGluIleLe 340
Db 13202 TCAGTTTCATCTGCTGCTGCTGCAAGTAACTTATATCCGACAGCTCTGGAATTTT 13261
QY 340 uThrAsnAlaThrGlnCythrLeuLeuCyAspProPheArgLeuCythrAspAsnG 360
Db 13262 AACAAACGCAACACAGTCACTTGTGTCTCTCCGACATATGCACTGATATATG 13321
QY 360 YlleMetIleAlaTrp----- 365
Db 13322 CATTAATGATTGATGTAAGCCACAGATATACGTCTTCACTCAATACTATGTAATAT 13381
QY 365 ----- 365
Db 13382 TAAATTCATTTTATCATACTAAGCCTTCTTCAGATCTTGAGCTATGATTTTAT 13441
QY 366 -----AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuH 379
Db 13442 TTTAATGCTTCTTATTAATGATGTAATGTAAGAACTAGCTGCTGGCATTTTAC 13501
QY 379 IAspIleGluGlyIleArgTyArgIleProIys 389
Db 13502 ATGACATGAAAGGATCCGCTATGAAACCAAG 13533

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:38:06 / Search time 18.9 Seconds
(without alignments)
2107.605 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILTKTAGVFPKSKRKRV.....DISKEVGSASIKVPLQKMEI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	634	29.8	463	2	probable O-sialoglycoprotein endopeptidase
2	553	26.0	387	2	probable O-sialoglycoprotein endopeptidase
3	548	25.8	365	2	probable O-sialoglycoprotein endopeptidase
4	548	25.8	366	2	probable O-sialoglycoprotein endopeptidase
5	540.5	25.4	344	2	probable O-sialoglycoprotein endopeptidase
6	539.5	25.4	359	2	probable O-sialoglycoprotein endopeptidase
7	535.5	25.2	367	2	probable O-sialoglycoprotein endopeptidase
8	524	24.7	421	2	probable O-sialoglycoprotein endopeptidase
9	516.5	24.3	335	2	probable O-sialoglycoprotein endopeptidase
10	492.5	23.2	341	2	probable O-sialoglycoprotein endopeptidase
11	488.5	23.0	337	2	probable O-sialoglycoprotein endopeptidase
12	488.5	23.0	337	2	probable O-sialoglycoprotein endopeptidase
13	484.5	22.8	337	1	probable O-sialoglycoprotein endopeptidase
14	484.5	22.8	337	2	probable O-sialoglycoprotein endopeptidase
15	483.5	22.8	337	2	probable O-sialoglycoprotein endopeptidase
16	475.5	22.4	325	2	probable O-sialoglycoprotein endopeptidase
17	475.5	22.4	342	2	probable O-sialoglycoprotein endopeptidase
18	469	22.1	340	2	probable O-sialoglycoprotein endopeptidase
19	467.5	22.0	354	2	probable O-sialoglycoprotein endopeptidase
20	465.5	21.9	346	2	probable O-sialoglycoprotein endopeptidase
21	465.5	21.9	354	2	probable O-sialoglycoprotein endopeptidase
22	463.5	21.8	343	2	probable O-sialoglycoprotein endopeptidase
23	460	21.6	336	2	probable O-sialoglycoprotein endopeptidase
24	459.5	21.6	346	2	probable O-sialoglycoprotein endopeptidase
25	457	21.5	327	2	probable O-sialoglycoprotein endopeptidase
26	455	21.4	336	2	probable O-sialoglycoprotein endopeptidase
27	455	21.4	348	2	probable O-sialoglycoprotein endopeptidase
28	451.5	21.2	344	2	probable O-sialoglycoprotein endopeptidase
29	439	20.7	344	2	probable O-sialoglycoprotein endopeptidase

30	439	20.7	346	2	probable O-sialoglycoprotein endopeptidase
31	437.5	20.6	412	2	probable O-sialoglycoprotein endopeptidase
32	431	20.3	348	2	probable O-sialoglycoprotein endopeptidase
33	430	20.2	341	2	probable O-sialoglycoprotein endopeptidase
34	409.5	19.3	338	2	probable O-sialoglycoprotein endopeptidase
35	409	19.2	352	2	probable O-sialoglycoprotein endopeptidase
36	406.5	19.1	344	2	probable O-sialoglycoprotein endopeptidase
37	406.5	19.1	344	2	probable O-sialoglycoprotein endopeptidase
38	405	18.8	374	2	probable O-sialoglycoprotein endopeptidase
39	400.5	18.8	346	2	probable O-sialoglycoprotein endopeptidase
40	400	18.8	336	2	probable O-sialoglycoprotein endopeptidase
41	391.5	18.4	335	2	probable O-sialoglycoprotein endopeptidase
42	379.5	17.9	344	2	probable O-sialoglycoprotein endopeptidase
43	361.5	17.0	407	2	probable O-sialoglycoprotein endopeptidase
44	361	17.0	302	2	probable O-sialoglycoprotein endopeptidase
45	353.5	16.6	315	2	probable O-sialoglycoprotein endopeptidase

ALIGNMENTS

RESULT 1
E84888
probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Jun-2003
C:Accession: E84888
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bente, M.I.; Town, C.D.; Fujii, C.Y.; B.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Gues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MIMD:20083487; PMID:10617197
A:Accession: E84888
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-463 <STO>
A:Cross-references: GB:AE002093; NID:G2583127; PIDN:AAB82636.1; GSPDB:GN00139
A:Gene: AtG945270
A:Map position: 2
A:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 29.8%; Score 634; DB 2; Length 463;
Best Local Similarity 37.9%; Pred. No. 2e-43;
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

QY	38	IVLGIFETSCDDPRAAAYVDSFGVYLGSAHISQTEVHLKTSQIVPPAQQLHRENIQRIYOE	97
DB	85	VVLGIFETSCDDPRAAAYVDSFGVYLGSAHISQTEVHLKTSQIVPPAQQLHRENIQRIYOE	141
QY	98	ALASAGVSPSDSAIATTKPGIALSLGVGLSPSLQVQLKKPFIPIHMEAHATIRL	157
DB	142	ALDKANITKEDKLSAVAVTIGPGLSLCLRVGRARARAFSLPIGVHMEHVALYL	201
QY	158	T-NKVEPPVLVILISGCHLALVQGVSDFLGLKSLDIAPEGMDLVKARSLIKHPBC	216
DB	202	VEQELSPFPMALLISGCHLALVLAHKLQYQLGTVDAIGAPKTAKLIDNH---	258
QY	217	STMSGKALIEHLAKQGNRHPDIKPELHAKNDPFPGLQHTYTDIKMKKEBEGIEKG	276
DB	259	RSQGPVAVELALDEGDASVKFNVMKHKDCNFSYAGLKTQVRLAIEK-----	308
QY	277	QLTSSADIAATVQHTMACHLVKTRTALIFCKORDLPPONNAVLAASGVASNFYIRA	336
DB	309	IRRADIASFORAVVHLSEKCRALIDMALE---LEPSIKMVTISGVASNKTVRLR	363
QY	337	LEILTNATQCTLLCPPEPLCTDNGIMIANNGIERLAGLILADIGIRYE-----	387
DB	364	LNINVENKTKLKVCPPEPLCTDNGVAVMTGLAEHFRVG-----RDPEPPATEPE	413
QY	388	PKCPFGVDSIKVEGEA 403	

Db 414 DYVYDLRRPMLGEBYAKGRSEA 436

RESULT 2

E71711

probable o-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii

C/Species: Rickettsia prowazekii

C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004

C/Accession: E71711

R/Anderson, S.G.E.; Zomrodipour, A.; Anderson, J.O.; Sichertz-Ponten, T.; Alsmark, U.

Nature 396, 133-140, 1998

A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A/Reference number: A71630; MUID:99039499; PMID:9823893

A/Accession: E71711

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-387 <AND>

A/Cross-references: UNIPROT:Q9ZEA8; GB:AJ235270; GB:AJ235269; NID:G3860572; PIDN:CAA1450

A/Experimental source: strain Madrid R

C/Genetics:

A/Gene: gcp; RP037

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 26.0%; Score 553; DB 2; Length 387;
Best Local Similarity 35.1%; Pred. No. 5 6e-37;
Matches 133; Conservative 56; Mismatches 130; Indels 60; Gaps 6;

QY 37 KIVIGIETSCDDTAAYVDEFGNVLGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRI 96

Db 2 KIIIGIESSCDDTAISITERRKILSNIIISQNTHEAVFGVVPRIARSHLSNLDQALK 61

QY 97 EALSASGVSPSDLSAATTIKPGIALSLGVGSLQVGVGKPPPIHMEAHALT 156

Db 62 NVLKKSNTLEISIAITSGPGLIGYIVGSMFARSLSSAKKPPFIATNHLBGHALTR 121

QY 157 LTNKVEPPLVLLISGGHCLLALVOGVSDPFLLGKSLDAPGMDLVKARSLIKHPRC 216

Db 122 LTNISVYLLLSGGHCGQFPAVAVGLGKRYKLTITDIDAVETDVKAKMNTL----- 175

QY 217 STWSGKAIEHLAKQGNRFHDIKPLHLAKNCDSPFTGLQHTVDKIIMK-KEKEGIRK 275

Db 176 -SFGGPEIEKRAKLGKPNHKYKFPKPIINSQNMNSFSGLKTAVRPLTLMNLEKVDV-- 232

QY 276 GQILSSADIAATVQHTMACHLVKRTTHALIFCKQ-----RD-- 312

Db 233 -----INDIASPQTIGAILSSKMDAIRLYKQILNDYEDINHPRTKLNKSFRKDEF 286

QY 313 -----LLPON-----NAVIVASGVASNFYIRALLETITNATQCTLLCP 351

Db 287 NKKLELECTRKRRIHIONSTRSNLNDITIVAGVAAANKYLQELISDCTRYGRLIAP 346

QY 352 PPRLCDNGIMIANMGIER 370

Db 347 PMHLCTDAAAMIAVAGLER 365

RESULT 3

O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AB2902

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L.

Emerge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCLell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: AB2902

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-365 <KUR>

A/Cross-references: GB:AE008668; PIDN:AA143632.1; PID:g17741154; GSPDB:GN00186

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Gene: gcp

A/Map position: circular chromosome

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 25.8%; Score 548; DB 2; Length 365;
Best Local Similarity 36.9%; Pred. No. 1.3e-36;
Matches 137; Conservative 58; Mismatches 142; Indels 34; Gaps 8;

QY 39 VLGIETSCDDTAAYV----DETGNVLGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRI 94

Db 7 ILGIETSCDETAASIVRRADRGELVSDVVLSQLSEHSYGGVVEIARAVAEALDTL 66

QY 95 VOBALSASGVSPSDLSAATTIKPGIALSLGVGSLQVGVGKPPPIHMEAHALT 154

Db 67 VESALDQAGVKLADVDAIATSGPGLIGLVLGIMTKAIAKAAKRPVYAINHLBGHALT 126

QY 155 IRLTNKVEPPLVLLISGGHCLLALVOGVSDPFLLGKSLDAPGMDLVKARSLIKHP 214

Db 127 ARLTDGLSPPYLMLLVSGHTQVLVVRGVEYERWGTITDIDALGEAPDKAKLGL-PPY 185

QY 215 ECSTWSGKAIEHLAKQGNRFHDIKPLHLAKNCDSPFTGL-----QHTVDKIIMKEKE 270

Db 186 -----GEPAVENAAKPDPRPPLPRPVGKARLDFSGGLKTAVRQAATALAPLSEQD 239

QY 271 EGIEKQILSSADIAATVQHTMACHLVKRTTHALIFCKQRLDPONNA--VLVASGVA 328

Db 240 -----IADICASFQKAVSRITKDRIGRLARFVE--FPHINGEPALVAVGVA 286

QY 329 SNFYIRALLETITNATQCTLLCPPRLCDNGIMIANMGIERLAKGILHIDEGIRYER 388

Db 287 ANQELRQTLQALCDTGFGRFVAPPHRLCTDAAAMIAWAGIERMAEG---ROADALEVAP 342

QY 389 KCPLGVDISKE 399

Db 343 RSRWPLDGSAB 353

RESULT 4

probable o-sialoglycoprotein endopeptidase (glycoproteinase) [imported] - Agrobacterium t

C/Species: Agrobacterium tumefaciens

C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C/Accession: D97677

R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume

A/Reference number: A97359; MUID:21608551; PMID:11743194

A/Accession: D97677

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-366 <KUR>

A/Cross-references: GB:AE007869; PIDN:AAK8373.1; PID:g15157858; GSPDB:GN00169

C/Genetics:

A/Gene: AGR C 4806

A/Map position: circular chromosome

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 25.8%; Score 548; DB 2; Length 366;
Best Local Similarity 36.9%; Pred. No. 1.3e-36;
Matches 137; Conservative 58; Mismatches 142; Indels 34; Gaps 8;

QY 39 VLGIETSCDDTAAYV----DETGNVLGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRI 94

Db 8 ILGIETSCDETAASIVRRADRGELVSDVVLSQLSEHSYGGVVEIARAVAEALDTL 67

QY 95 VOBALSASGVSPSDLSAATTIKPGIALSLGVGSLQVGVGKPPPIHMEAHALT 154

Db 68 VESALDQAGVKLADVDAIATSGPGLIGLVLGIMTKAIAKAAKRPVYAINHLBGHALT 127

```

Oy 155 |||TNVPEPEVLIIIVSGHCLLIVGVSOFLLGKSLDIAKMDLDKVARELSLKH 214
Db 128 ARITDGSFPEVLMIVSGHTQVLVKGVSBERMGTITDDBAGFADTKALGL- -PY 186
Oy 215 ECSTWSGKAIEHLAKQGNREHPDIKEPLHAKXNDFSEYGL- -QHVYDKIMKKEKE 270
Db 187 - - - - -GSPAVENAAKGDPRFPPLPRMVEARLDPSFGKTAVRQATATAIAPSEQD 240
Oy 271 EGIEKQIILSSADIAITVQHTMACHLVKRTIRALLFCGRDILLPOMNA- -VLVAGSGVA 328
Db 241 - - - - -IADICASFQKAVSERTLDRIGRGIAEFVE- -FPHINSEPALVAVAGVA 287
Oy 329 SNFYIRRALEIILTNATQCTLLCPPEPLCTDNGIMIAWNGIERLRAGILHDIIEGIRYEP 388
Db 288 ANQEIHQITQALCDTHGRFVAFPRHRLCTDNAAAMIAWGLEMAEG- - - - -RQADALEVAR 343
Oy 389 KCEPLGVDSIKE 399
Db 344 RSRWPLDGSAE 354

```

```

RESULT 5
E97707
O:siatloglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Rickettsia conorii (strain
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: E97707
R:Ogata, H.; Audic, S.; Renesto-Audifferen, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700, MUID:21442074; PMID:11557893
A:Accession: E97707
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-344 <KUR>
A:Cross-references: GB:AE006914; PIDN:AL0259.1; PID:g15619097; GSPDB:GN00173
C:Genetics:
A:Gene: gcp
C:Superfamily: O-siatloglycoprotein endopeptidase
C:Keywords: hydrolase; metalloproteinase

```

[illegible]

O-sialoglycoprotein endopeptidase (EC 3.4.24.57) (imported) - *Brucella melitensis* (strain
C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #ext_change 15-Feb-2002
C:Accession: AB3274
R:DelVecchio, V.G.; Kaparatel, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A:Reference number: A03252; PMID:11756688
A:Accession: AB3274
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-359 <KUR>
A:Cross-References: GB:AE008917, PIDN:AAL51357.1, PID:g17982056, GSPDB:GNO0190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0175
A:Map position: I
C:Superfamily: O-sialoglycoprotein endopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match	25.4%	Score 539.5	DB 2	Length 359
Best Local Similarity	38.6%	Pred. 6.3e-36		
Matches 132	Conservative 52	Mismatches 127	Indels 31	Gaps 7
Qy	39	VLGIGTSCDDPAAAVVDE----	TGNVLGEALHSOTVEVLKTGVI	PPAAQOLHRENTORI 94
		: : :	: : :	: : :
Db	3	VLGIGTSCDDPAAAVVDE----	VERDMDGEGRIILSNVLSQIAHEP	PGGVPELIAAQAHEALDRL 62
Qy	95	VOEALASGVSPDLSAIAITTIKQ	LALSLGVLSFSLQVLGQLKKEPI	PIHMEAAHLT 154
		: : :	: : :	: : :
Db	63	VDRLANDAGLKLIEYDVAATA	GAGLIGLIGLIGLMTATALMAAQ	KPEYAVNHHGHALT 122
Qy	155	IRLNNKVFPPRLVILISGCHGL	LALNVGVSPFLILKSLDILAR	GDMLDKVARLSLIGHP 214
		: : :	: : :	: : :
Db	123	ARLTIDGLFPPLLLLVSGH	QOMVLVREIGDYERIGITTD	IALGBAFKTKLGL-PPY 181
Qy	215	ECSTMSGKALEHLAKOGN	PHFPIKEPLHAKKNCDSFSTG	-----OHWDTIKMKKEK 269
		: : :	: : :	: : :
Db	182	-----GGPAVERNALQGD	QKRFALPRPLKEALDLFS	GLKTRAVRQTALVELVPLTDQ- 234
Qy	270	BEGIEKQILISSADIAATV	QHTACHLVKXTRHALLFC	KORDLLPQ--NNAVLVASGVA 328
		: : :	: : :	: : :
Db	235	-----DVIDICASFOA	VADTLDSDRVSLSERFKTE--	PPDCATPSLVAGVA 281
Qy	329	SNFYIRALLETILTA	QCTLLCAPPRLCTDNGIM	IANGIER 370
		: : :	: : :	: : :
Db	282	ANKTLRALLENILCTR	HGFAPLAPLANTCTD	NAAMIANAAGAR 323

RESULT 7
P87257
peptidase M22 family protein [imported] - *Caulobacter crescentus*
C|Species: *Caulobacter crescentus*
C|Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C|Accession: P87257
R|Nieman, W.C.; Feldbylm, T.V.; Paulsen, I.T.; Nelson, K.R.; Eisen, J.; Heidelberg, J.P.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.B.; Haft, D.H.; Kolton
n, J.; Brimacombe, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4144, 2001
A|Title: Complete Genome Sequence of *Caulobacter crescentus*.
A|Reference number: A87249; MUID:21173698; PMID:11259647
A|Accession: P87257
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-367 <STO>
A|Cross-references: GB:AB005673; NID:g13421168; P1DN:AAK22058.1; GSPDB:GN00148
C|Genetic8:
A|Gene: CC0071
A|Superfamily: O-sialoglycoprotein endopeptidase

RESULT 6
AB3274

Query Match	25.2%	Score 535.5;	DB 2;	Length 367;
Best Local Similarity	37.9%	Pred. No. 1.4e-35;		

Matches 136; Conservative 54; Mismatches 144; Indels 25; Gaps 7;

QY 38 IVLGIEISCDTAAAVV---DEGNVIGEAHISQTEVHLKTGGIVPPAAQOLHRENIOR 93
 Db 10 LILGIEISCDTAAASVVRRAADGIVTVSSVIGTFEKKIAPPGVVPETIAAAHVESIDA 69
 QY 94 IVOEASASGVSPDLSAATITIKPGLALSIGVGLSFSIQVGOLKKPPIPIHMEAHAL 153
 Db 70 IAAEVRRAAGVFGDLDDVAANAGPGLVGVVWGLAFGCAVALARGAPLVANHLEGHAV 129
 QY 154 TIRLTNNKVPFPLVLLISGGHCLALVGVSDPFLIGSLDIAPDMDLKVARRLSLIGH 213
 Db 130 SARLGADIAVPFPLLVVSGHQLLEVSGVCKKLGITTDAAAEAFDKIKSLGL-PY 188
 QY 214 PECSMGSGKAIENHAKGNRHFPIKPRPLHAKKCDSPFGIQVNTDKIIMKKKEBEI 273
 Db 189 P-----GGPALKEKLVAGSDPFRVALPRALGKRCDCSPSLKTPAAAIATLTDD-- 240
 QY 274 EKGQILSSAADIATVQHTMACHLVKRTTRAILPCKORDLIPQNNAVLVASGVASNFYI 333
 Db 241 -----ARRDLAAGVQAIAIARQLSERVDRAMKLYK--DSHPEDLRFVAVAGVANAAGV 291
 QY 334 RALEIILTNATQCTLLCPPEPLCTDNGIMIANNGIERLAGIGLHDIEGIRYBPCPL 392
 Db 292 RAALLADCEKKGFSBAPAPPLACTDNAAIMIALAGARL--ALGIFDDLDAIA-RPRWPL 347

RESULT 8
 T18825
 hypothetical protein C01G10.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T18825
 R:Matthews, L.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19027
 A:Accession: T18825
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-421 <WILL>
 A:Cross-references: EMBL:Z81030; PIDN:CA802716.1; GSPDB:GN00023; CESP:C01G10.10
 A:Experimental source: clone C01G10
 C:Genetics:
 A:Gene: CESP:C01G10.10
 A:Map position: 5
 A:Initrions: 31/2; 122/2; 177/1; 272/3; 315/1; 353/2
 C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.7%; Score 524; DB 2; Length 421;
 Best Local Similarity 33.4%; Pred. No. 1,4e-34;
 Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;

QY 39 VLGIETSCDDTAAAVVDEGNVIGEAHISQTEVHLKTGGIVPPAAQOLHRENIORIVQEA 98
 Db 25 VLGIETSCDDTAAAVVDEGNVIGEAHISQTEVHLKTGGIVPPAAQOLHRENIORIVQEA 98
 QY 99 LSASGVSPDLSAATITIKPGLALSIGVGLSFSIQVGOLKKPPIPIHMEAHALTRILT 158
 Db 85 LNDAGTSPKDLDAVAVTVPGVIALKEGISAIAIFAKKGRPLPLPVHMEAHALSTILV 144
 QY 159 -NKVEFPPLVLLISGGHCLALVGVSDPFLIGSLDIAPDMDLKVARRLSLIGHREPS 217
 Db 145 DDSVRFPPSAVLLISGGHCLALSVADEVEKFKLVGVSGSPGECIDKVAQGLDL-GSEFD 203
 QY 218 TMSGGKALEHIAKOGN---RFHFDIKPPLHAKNCDPSFTGLQHTVDKIMKKKEEGIE 274
 Db 204 GIVHGAATEIILASRSASGCHLRYPFLPVPKANNFQDIKSYLMLERLAKNSTSID 263
 QY 275 KGQILSSADIAATVQHTMACHLVKRTTRAILPCKORDLIPQNNAVLVASGVASNFYIR 334
 Db 264 -----IPDFCASIQNTVAHRISSKMHFPESLSBQELPQO---LVIGGVAAANOYIF 313
 QY 335 RALEIILTNATQCTLLCPPEPLCTDNGIMIANNGIERLAGIGLHDIEGIRYB----- 388

Db 314 GAISKLSAAHNVYTIKVLISLCTDNAMENIAYSGL-----LMLVNRSEAIWPRPNIDPPT 367
 QY 389 -----KCPFGVDISKEVGASIKVPQLKM 412
 Db 368 IYAHARSIDIGTDASSSI-----IDTPRRKL 392

RESULT 9
 G70369
 sialoglycoproteinase - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
 C:Accession: G70369
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
 V.
 Nature 392, 353-358, 1998
 A>Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666; PMID:9537320
 A:Accession: G70369
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-335 <AOQ>
 A:Cross-references: UNIPROT:O66986; GB:AE000708; NID:G2983356; PIDN:AA06951.1; PID:G2983
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: gcp
 C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.3%; Score 516.5; DB 2; Length 335;
 Best Local Similarity 36.5%; Pred. No. 4.2e-34;
 Matches 130; Conservative 66; Mismatches 125; Indels 35; Gaps 10;

QY 40 LGIETSCDDTAAAVVDEGNVIGEAHISQTEVHLKTGGIVPPAAQOLHRENIORIVQEA 99
 Db 4 LAVETSCDETALAIYDDQKGVGNVILSQAVHSPGVPPEISAREHTRNIIPIFDRL 63
 QY 100 SASGVSPDLSAATITIKPGLALSIGVGLSFSIQVGOLKKPPIPIHMEAHALTRILT 159
 Db 64 KESRINLEIDIFISFLTLEGLISLVGAFAALAEYERKPLVPHHLEGHYSVFLK 123
 QY 160 KYEPFPLVLLISGGHCLALVGVSDPFLIGSLDIAPDMDLKVARRLSLIGHREPS 219
 Db 124 KYEPFPLVLLISGGHCLALVGVSDPFLIGSLDIAPDMDLKVARRLSLIGHREPS 219
 QY 220 SGGKALEHIAKOGNRFHFDIKPPLHAKNCDPSFTGLQHTVDKIMKKKEEGIEKGQIL 279
 Db 178 -GGPIIDRLAKGKGL-YPLPKPLMEEGNINFSFGLK---TALIMLKKKKVVRK---- 228
 QY 280 SSAADIAATVQHTMACHLVKRTTRAILPCKORDLIPQNNAVLVASGVASNFYIRALEI 339
 Db 229 ---EDIAVSFOETVEVILLES---LWMMKKTGIR---LVVGVGSANSRLR---EV 274
 QY 340 LTNATQ---CTLLCPPEPLCTDNGIMIANNGIERLAGIGLHDIEGIRYBPCPL 392
 Db 275 FKXASQYGFELIYIPHSISTDNALMIAVAGHFRKGVAVAPLDVNP---QPMIPL 327

RESULT 10
 H83572
 O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: H83572
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brj
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: H83572
 A:Status: preliminary
 A:Molecule type: DNA

Db 230 --OTRADIAAEEDAVVDTLAKSKRA-----LDQTFGRVLVIAGGSANQTLRLK 278
 Qy 337 LEILTNATQCTLLCPPEPRCTDNGIMIANGIERLBAHGLHDIGIRYEPKPL 392
 Db 279 LADMMQKRGGEVFTYARPEFCTDNGAMTAYAGWRLRSN--LNSELTSVSVPRWPL 331

Search completed: February 16, 2005, 13:09:20
 job time : 19.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:19:07 ; Search time 4918.95 Seconds
(without alignments)
3203.653 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 2125
Sequence: 1 MLILTKTAGVFFKPSKRVY.....DISKEVGASIKVPLKMEI 414

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 1902134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+g2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US10649273/runat_14022005_114703_16411/app_query.fasta_1.1429
-DB=EST -QPMT=fastcap -SUFFIX=rcs -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=50
-UNITS=dits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273.QCGN_1_1_6799.Qrunat_14022005_114703_16411 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOD
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1827	86.0	2284	3	AK045669 Mus muscu
2	1819	85.6	1622	3	AK011265 Mus muscu
3	1634	76.9	1609	3	BC030671 Mus muscu
4	1447	68.1	870	5	BQ423651 AGENCOURT
5	1353.5	63.7	852	5	BX391919 BX391919
6	1279	60.2	922	5	BQ961028 AGENCOURT
7	1173	55.2	701	2	BR740611 601595739
8	1158	54.5	749	7	CR982692 4115331 B
9	1144	53.8	800	7	CO738006 SLH03C19

10	1132	53.3	1109	5	BM907988	BM907988 AGENCOURT
11	1111	52.3	765	7	CF114247	CF114247 Shultzomi
12	1097	51.6	640	5	B0636028	B0636028 hdd0d11.y
13	1045	49.2	730	7	CN823245	CN823245 Oa. gplbn
14	1038.5	48.9	661	7	CK833139	CK833139 4057170 B
15	1018.5	47.9	879	5	BQ431135	BQ431135 AGENCOURT
16	1017	47.9	1173	6	CD508917	CD508917 CDA93-E05
17	997	46.9	723	7	CK942291	CK942291 4065791 B
18	978	46.0	579	7	CK819035	CK819035 if06d05.y
19	966	45.5	822	7	CF257246	CF257246 pha008.gu
20	963	45.3	577	7	CV028547	CV028547 7089 Full
21	954.5	44.9	808	7	CK792978	CK792978 AGENCOURT
22	954	44.9	833	5	BUI59877	BUI59877 AGENCOURT
23	951.5	44.8	728	4	BG46856	BG46856 60248845
24	935.5	44.0	879	5	BU256052	BU256052 603745167
25	935	44.0	871	2	BF308169	BF308169 601887411
26	925	43.5	658	7	CF362328	CF362328 829596 MA
27	923	43.4	811	7	CN232675	CN232675 WLB051B07
28	920	43.3	859	5	BU292595	BU292595 603604785
29	916	43.1	831	7	CK455183	CK455183 915735 MA
30	915	43.1	812	5	BU246489	BU246489 603784202
31	912	42.9	878	7	CK456228	CK456228 916887 MA
32	911	42.9	661	7	CF787040	CF787040 854903 MA
33	907	42.7	717	7	CK459683	CK459683 929895 MA
34	906	42.6	597	6	CB272391	CB272391 ma157512.
35	902	42.4	533	4	BM126453	BM126453 if06d05.y
36	893	42.0	574	5	BQ130908	BQ130908 if06d05.y
37	893	42.0	687	7	CF360919	CF360919 827217 MA
38	879	41.4	737	7	CV073257	CV073257 AGENCOURT
39	867	40.8	616	6	CB438468	CB438468 687077 MA
40	857.5	40.4	789	5	BU242187	BU242187 603781023
41	838	39.4	637	7	CK941819	CK941819 4065407 B
42	831	39.1	883	7	CR444993	CR444993 CR444993
43	825.5	38.8	792	4	BG333362	BG333362 602431206
44	819.5	38.6	792	5	BU403563	BU403563 604138456
45	819	38.5	668	5	BU285782	BU285782 604166893

ALIGNMENTS

RESULT 1
AK045669
LOCUS
DEFINITION
AK045669 2284 bp mRNA linear HTC 03-Apr-2004
Mus musculus adult male corpora quadrigemina cDNA, RIKEN
Full-length enriched library, clone:B230219017 Product:similar to
PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo sapiens), full insert
sequence.

ACCESSION
AK045669.1 GI:26337528
VERSION
AK045669.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
REFERENCE
AUTHORS

TITLE
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159

REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

TITLE	Pujiwke,S., Inoue,K., Togawa,Y., Irawa,M., Ohara,B., Matsubiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format
MEDLINE	sequencing pipeline with 384 multiplexillary sequencer
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE	20530913
AUTHORS	11076861
TITLE	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
JOURNAL	Functional annotation of a full-length mouse cDNA collection
REFERENCE	Nature 409, 685-690 (2001)
AUTHORS	5
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
REFERENCE	Nature 420, 563-573 (2002)
AUTHORS	6 (bases 1 to 2284)
TITLE	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Horii,F., Imomani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohashi,N., Okazaki,Y., Saito,R., Saitoh,K., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Saeki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takai-Akahira,S., Tanaka,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
JOURNAL	Direct Submission
REFERENCE	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
AUTHORS	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
TITLE	Please visit our web site for further details.
JOURNAL	URL: http://genome.gsc.riken.jp/
REFERENCE	URL: http://lennon.gsc.riken.jp/.
AUTHORS	Location/Qualifiers
TITLE	1. .2284
JOURNAL	/organism="Mus musculus"
REFERENCE	/mol_type="mRNA"
AUTHORS	/strain="C57BL/6J"
TITLE	/db_xref="PANTOM,DB:B230219017"
JOURNAL	/db_xref="taxon:10090"
REFERENCE	/clone="B230219017"
AUTHORS	/sex="male"
TITLE	/tissue_type="corpora quadrigemina"
JOURNAL	/clone_lib="RIKEN full-length enriched mouse cDNA library"
REFERENCE	/dev_stage="adult"
AUTHORS	221. .1465
TITLE	/note="unnamed protein product; putative
JOURNAL	similar to PUTATIVE SIMILGUCROPROTEIN TYPE 2 (Homo sapiens) (SPTR Q9H4B0, evidence: FASRY, 80%ID, 100length, match=1242)"
REFERENCE	/codon_start=1
AUTHORS	/protein_id="BAC32450.1"
TITLE	/db_xref="GI:26337529"
JOURNAL	/translation="MTLRLRAGAIIPKPKSKVGLRPSVHPRTLSCHLVGIETSCDRTGAAYVDENGVLRGALRSOTVNRKGVLPVPAQVLENIORVETLSAOCRTIPSDLSAIAITTKRGLALSLGVLGSLVQVQFKKPIPIHMEAHALITLTKRVSFPPVLLISGCHLALVQGVSDPFLILGSLDIAGDMLDKVARSLIKPECSYVMSGKLEIOLAKGNRPHFTINPMQNAKCDPFTGIQHTDITLTHKEKEGIEKGK

		polya_signal		OIISSADIIAAVOHATGACIKAKTRHAIIFCQKINLSPANNVLVSGVANSLYTR	
		polya_site		KALEIVANATQCTLCPPPLCTCDNGNIMIMANIGERLRAGLGVLDHVEDIRYBKCPCL	
		/note="putative"		GVDISREVARAIIVPRILKAL"	
		2262..2267			
		2284			
		/note="putative"			
ALIGNMENT SCORES:					
Pred. No.:		5.12e-194	Length:	2284	
Score:		1827.00	Matches:	352	
Percent Similarity:		90.58%	Conservative:	23	
Best Local Similarity:		85.02%	Mismatches:	39	
Query Match:		85.98%	Indels:	0	
DB:		3	Gaps:	0	
US-10-649-273-2 (1-414) x AK045669 (1-2284)					
QY	1	MetLeuIleLeuThrLysThrAlaGlyValPhe	helyProSerLysAlaGlyValTyr	20	
DB	221	ATGCTATGTTAAGAAAGACAGCAGGAGCTATTCCCAAGCCCAAGAGTAAAGTTAT		280	
QY	21	GlulPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu		40	
DB	281	GGATTTTAAACAAGATTTATGTCTTCATCCCAAGCTCTCTTGTCAATAACTGGTCTG		340	
QY	41	GlyIleGluThrSerCysAspAspThrAlaAlaValAlaSerGluThrGlyAsnVal		60	
DB	341	GGATTGTAAACCACTGTGATGACACAGGACGGCTGTGGTGAATGAACCTGGAAATGTG		400	
QY	61	LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro		80	
DB	401	CTGGGGAGACACCTGCACTCCCAAACTCAGGTCATCTGAAMACAGGTGGATTGTTCT		460	
QY	81	ProAlaIaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer		100	
DB	461	CCAGTAGCTACACAACTTCACAGAAAAATTTTAAAGAAATGATGAAGAAACCTTTTCT		520	
QY	101	AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu		120	
DB	521	GCCTGTAGAATCACCCCAAGGACTCTCAGCAATTGCCAATCAACCAACCGGAGCTG		580	
QY	121	AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys		140	
DB	581	GCCCTAAGCTTGAGAGTGGCTTATCCTTTAGCTTACAGCTAGTAAATCAGTTAAAAAG		640	
QY	141	ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys		160	
DB	641	CCATTTCATCCCGAATTCATCATAGGAGGCTCACCGACTGCTAATTAGCTTCACCAATAAA		700	
QY	161	ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal		180	
DB	701	GTAAGATTTCTTTTATTTAGTTCTTTTATTTCTGGCGGTCACTGCTGTGGCAATTAGTC		760	
QY	181	GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet		200	
DB	761	CAAGGTGTTTCCGATTTCTGCTCTCTTGGAAGCTTTTGACATTAAGCACCGGCAATG		820	
QY	201	LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer		220	
DB	821	CTTGACAAAGGTGGCAAGAACTTTCTTTATCAAAACATCCAGATATGTTCTACAAATGAGT		880	
QY	221	GlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLys		240	
DB	881	GGTGAAGAAAGCTATTAAGAACAGTTGGCCAAACAGGAATATGATTTCAATTTACTATCAAT		940	
QY	241	ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr		260	
DB	941	CCACTATGCAGAGATGCTTAAGAAATTTGGGATTTTCTTTCAAGGGACTTCAACATATTAAT		1000	
QY	261	AspLysIleIleLeuLysLysGlnLysGlnGlyIleGluLysGlyGlnIleLeuSer		280	

[illegible]

TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 1622)
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukushima, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirotsuka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komou, H., Kouda, M., Koya, S., Kuwahara, C., Matsuyama, T., Miyazaki, A., Nishik, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawara, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyai, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submision
JOURNAL	Submitted (10-JULY-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.jp/) for further details.
FEATURES	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGACGAGAAGATCCAGACTCTTTTTCCTTTTTCCTTA 3'], cDNA was prepared by using triazole thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGGATCTCGATTAAATTAATGATCCCCCCC 3']. cDNA was cleaved with XhoI and SacI. Cloning sites, 5' end: XhoI; 3' end: SacI. Host: SOLR.
source	Location/qualifiers 1. 1622 /organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="PANTOM_DB:2610001M19" /db_xref="taxon:10090" /clone="2610001M19" /tissue type="whole body" /clone lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days embryo" 207. 1451 /note="unnamed protein product; putative similar to PUTATIVE SIMALOGICPROTEASE TYPE 2 (Homo sapiens) (SPTR Q9HA80, evidence: FASTX, 80%ID, 100%length, match=1242)" /cdon_start=1 /protein_id="BAB27506.1" /db_xref="GI:12847276" /translation="MLMTARTTAGAIPKPKSKVYGFLRRFSVHPRTLSCHLVGIFIEISCDPDGAAYVDERTGVNGLBALHSOTVHALTKGGIVNPVAOOLHEENTOIYEETLSLRKRIPSDLSAATITIGIALSLCYGSFSLQVNOPKKPRPIHHMEAHALLTLRTLTKRVEPPVLILSGCHLALVOGYSDPLILGKSLDIAGMDLDKARALSILKHPECTMSGSAIBOLAKODGNRPHTINPMONARKDSCFTSLQIHDTLLITKEKESEBIEKEWISGASIDIAAVOHATACHAKTRTHAPITFCOKNLISPNANVTLVSGVASNTLYIRVLESTVRAAFQCTLCFFPTLTCDTGIMIAMNGIERLRAGIAGVLAHVEDIRYREKCPCLGVDSIREVAALAIKVRLKNAL"
CDS	polya_signal 1605..1610 /note="putative" polya_site 1622

ORIGIN /note="putative"

Alignment Scores:

Score: 2.38e-193 Length: 1622
Percent Similarity: 1819.00 Matches: 351
Best Local Similarity: 90.34% Conservative: 23
Query Match: 84.78% Mismatches: 40
Gaps: 0
Indels: 0

US-10-649-273-2 (1-414) x AK011265 (1-1622)

```

Qy 1 MetLeuLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db 207 ATGCTTAATGTTAAAGAGACAGACGAGGAGTATCCCAAGCCCCCAAGAGTAAGTTAT 266
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLysPheLeuHisLysL1LeuValLeu 40
Db 267 GGATTTTAAAGAAATTAGTGTTCATCCAGAACTCTCTCTTCATTAATGAGTCTCG 326
Qy 41 GlyL1LeuThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
Db 327 GGAATTGAAACCAAGCTGTGATGACACAGAGCGCTGTGTGTGATGAAACCTGGGAATGTG 386
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLysLysThrGlyL1LeuValPro 80
Db 387 CTGGGGGAGAGACATGCACTCCCAACTCAGGTTCTATGAAACAGTGGAGTGTCTCT 446
Qy 81 ProL1aIaGlnGlnLeuHisArgGlyAsnL1GlnArgL1LeuValGlnGluL1aLeuSer 100
Db 447 CCACTAGCTCAACAACCTTCAAGAAAATATTCACAGAAATGTAAGAAATCTTTCT 506
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrL1LeuSerProGlyLeu 120
Db 507 GCCTGTGAATCACCCCAAGCATCTTCAGCAATTCGCAATCAATCAACCGGAGACTG 566
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 567 GCCCTAAGCTGGAGAGTGGCTTATCTTACCTTAACAGTGAATCAATGTTTAAAAAG 626
Qy 141 ProPheLeuProIleHisIleMetGlyAlaHisAlaLeuThrL1ArgLysThrAsnLys 160
Db 627 CCATTCACTCCGATTCATCAATGAGGCTCAGCATGACTATTAAGGCTCCATTAATA 686
Qy 161 ValGluPheProPheLeuValLeuLeuL1SerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 687 GTAGAATTTCTTTTGTAGTTCTTTGATTCTGGCGGTCATGCTGTTGGCATTTAGTC 746
Qy 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspL1LeuAlaProGlyAspMet 200
Db 747 CAAAGTGTTCCTGATTCCTGCTCTTGGAGATCTTGGACATGACACAGGCAATG 806
Qy 201 LeuAspLysValAlaArgArgLeuSerLeuL1LysHisProGlyCysSerThrMetSer 220
Db 807 CTTAACAAAGGTGGCAAGAACTTTCTTAATCAACATCCAAATGTTTCTCAATGAGT 866
Qy 221 GlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisAspLeuL1Lys 240
Db 867 GGTGAAAAGCTATAGAACAGTGGCCAAAGACGGAATAGATTCATTTTCAATCAT 926
Qy 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 927 CCACCTTGAAGCAAGTCAAGAAATTTCTTCTTTCACAGGACTTCACAAATATTTACT 986
Qy 261 AspLysL1IleMetLysLysGlyLysGlyGlnGlyL1LeuGlyGlyGlnL1LeuSer 280
Db 987 GATTAAGCTAATTAACACACAGGAAAGAAAGAGGCAATTAAGAGAGCAATTCCTGTA 1046
Qy 281 SerAlaIaAspL1LeuAlaIleThrValGlnHisThrMetAlaCysHisLysValLysArg 300
Db 1047 TCACTGAGACATTTGCTGCTGCGGTACAGCATGACAGCGCTGCACTTCGGAAGAA 1106

```

```

Qy 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
Db 1107 ACACATCGCGCATATTCGTTTGGACAGAGAAAATTTCTCTCCAGCTAACGACGTA 1166
Qy 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrL1ArgArgAlaLeuGluL1Leu 340
Db 1167 TTAGTGTATCTCGAGAGTGTTCAGAGTAACTGTATCATCCGAAMACATTTGAAATTTGTC 1226
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuLeuSerThrAspAsnGly 360
Db 1227 GCAAAATGCAACGACAGCGCTGTGTGTCCACTCCCAAGACTGTGCAATGAGC 1286
Qy 361 IleMetL1LeuAlaTrpAsnGlyL1GlnArgLeuArgAlaGlyLeuGlyL1LeuHisAsp 380
Db 1287 ATCAATGATTCATGGAATGGAATGGAATTAACGCTGCTGGCGCTTTACATGAT 1346
Qy 381 I1eGlyGlyL1ArgTrpGluProLysCysProLeuGlyValAspL1LeuSerGlyVal 400
Db 1347 GTAGAAGACATCCGATGATGAACCAAAATGCTCTTGGAGTGAACATTCAGAGAAATT 1406
Qy 401 GlyGluAlaSerL1LysValProGlnLeuLysMetGluL1Leu 414
Db 1407 GCAGAACTGCTCAATAAAGTCCGCAATTAATAAATGCACTT 1448

```

RESULT 3
BC030671
LOCUS
DEFINITION
MUS MUSCULUS O-61aIoglycoprotein endopeptidase-like 1, mRNA (CDNA clone IMAGE:1226118), containing frame-shift errors.
ACCESSION
BC030671
VERSION
BC030671.1 GI:21040459
KEYWORDS
HTC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1609)
Strusberg, R.L., Pelting, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Schmen, C.M., Schaller, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, D., Hsieh, F., Diatchenko, L., Marsina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Ueda, T.B., Toshitaka, S., Carninci, P., Prange, C., Kana, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huijck, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywicki, M.I., Skalska, U., Smallus, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL
MEDLINE
PUBMED
22388257
12477932
REFERENCE
2 (bases 1 to 1609)
Strusberg, R.
Direct Submission
Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIG-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadan@systemsbio.org
 Anup Madan, Jessica Fahay, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
 Series: IRAX Plate: 66 Row: e Column: 10
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1..1609
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1226118"
 /issue_type="Thymus gland, mouse"
 /clone_lib="Soares_thymus_2mbmt"
 /lab_host="DH10B"
 /note="Vector: pT7T3-Pac"

ORIGIN

Alignment Scores:

Pred. No.: 1.64e-172 Length: 1609
 Score: 1634.00 Matches: 327
 Percent Similarity: 83.86% Conservative: 21
 Best Local Similarity: 78.80% Mismatches: 40
 Query Match: 76.89% Indels: 27
 Gaps: 2

US-10-649-273-2 (1-414) x BC030671 (1-1609)

QY 1 MetheulleuThrlYThrAlaGlyValPhePheProSerIySArgIyValTyr 20
 Db 229 ATGTATATGTTAAAGAGACAGACAGACTATCCCAAGCCCAAGATTAAT 288
 QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisIleValLeu 40
 Db 289 GGATTTTAAAGAAATTAAGTTCATCCCAACTCTCTGTGCATTAAGTGTCTCG 348
 QY 41 GlyIleGluThrSerGlyAspAspThrAlaAlaValIleArgIleThrGlyVal 60
 Db 349 GGAATTGAACACACTGTGATGACACAGACCGCTGTGTGATGAAGAACTGGGAATGG 408
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeu-LysThrGlyIleValPr 80
 Db 409 CTGGGGAGACACTGCACTCCCAACTCAGGTTCTGAAAGT-----GGATTTGTC 462
 QY 80 OProAlaAlaGlnIleuHisIleArgIleValAsnIleGlnArgIleValGlnIleValLeu 100
 Db 463 TCAGTAGTCTCAACACTTCACAGAGAAATATTCAGAAATAGTAAGAAACTTTTC 522
 QY 100 rAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLe 120
 Db 523 TGCTGTAGATATCAACCCCAACGATCTCTCAGCAATTCACATCAACCCGGAAGT 582
 QY 120 uAlaLeuSerIleuGlyValGlyLeuSerPheSerIleuValGlyGlnLeuVal 140
 Db 583 GGCCCTAAGCTTGAGAGTGGCTTATCTTATGCTTACAGCTAGAAATCAGTTTAAAAA 642
 QY 140 sProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIly 160
 Db 643 GCATTCATCCCGATTATCATCAATGAGGCTCAAGCACTGATCTATAGCTCAACCAATTA 702
 QY 160 sValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlyLeuLeuAlaLeuVal 180
 Db 703 AGTAGAATTCCTTTTAACTCTTCTTCAATTCGCGGCTCAAGCTGATGCAATTAAGT 762
 QY 180 lGlnGlyValSerAspPheLeuLeuLeuGlyLysSerIleuAspIleAlaProGlyAspMe 200

Db 763 CCAAGGTTTCCGATTTCTGCTCTGGAGATCTTTGGACATGACACAGCGCAAT 822
 QY 200 tLeuAspIleValAlaIleArgIleSerLeuIleLysIleAspProGluCysSerThrMetSe 220
 Db 823 GCTTGCAAGGTGGCAAGAACTTTCTTTAAACAACATCCAGAAATGTTCAATGAG 882
 QY 220 rGlyGlyLysAlaIleGluHisIleuAlaLysGlnIleAsnArgPheHisPheAspIleu 240
 Db 883 TGGTGAAGAAAGCATATGACAGATGGCCAAAGACGAATATGATTCATTTACTATCA 942
 QY 240 sProPheLeuHisIleAlaLysAsnGlyAspAspPheSerPheThrGlyLeuGlnHisVal 260
 Db 943 TCACCTATGACAGAAATGATTAAGATTCGATTTTCTTTCACGGGACTTCACATATTAC 1002
 QY 260 rAspIleIleIleMetLysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSe 280
 Db 1003 TGAATAGCTAATTAACACACAGAGAAAGAAAGACATTAAGAGGGCAATCTGTGC 1062
 QY 280 rSerAlaAlaAspIleAlaAlaIleThrValGlnHisThrMetAlaCysHisIleuValIySar 300
 Db 1063 ATCAGCTGCAGCATTTGCTGCTCGGTACAGCATGCAACAGCTGCCACTTGCAGAAAG 1122
 QY 300 gThrHisIleAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaIyA 320
 Db 1123 AACACATCGCGCTATTCTGTTTTCAGAGCAAGAAATTTGCTCTCCAGCTAACCCAGT 1182
 QY 320 lLeuValAlaSerGlyGlyValAlaSerAsnPhenIleIleArgAlaLeuGluIleLe 340
 Db 1183 ATTAGTGTATCTGGAGGTTGTCAGATTAATGTTACATCCGAAAGCATTTGAAATTTGT 1242
 QY 340 uThrAsnAlaThrGlnCysThrLeuLeuLeuCysProProArgLysLeuGlyThrAspAnG 360
 Db 1243 CGCAATGCAAGCGACATGACGTTGTGTGTCCACTCCAGACTGTGACATGACATGG 1302
 QY 360 yIleMetIleAlaIleTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAs 380
 Db 1303 CATCATGATTGCA-----TGATGTCTCTTGGAGATGATATTCAGAGAAATG 1315
 QY 380 pIleGluGlyIleArgTyrluProluLysCysProLeuGlyValAspIleSerLysGluVal 400
 Db 1316 -----TGATGTCTCTTGGAGATGATATTCAGAGAAATG 1350
 QY 400 lGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
 Db 1351 TGCAGAACTGCCATTAAGATCCGCAATTAAGATGCACTT 1393

RESULT 4
 B0423651
 LOCUS
 DEFINITION
 AGENCOURT_7790948 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065828
 5', mRNA sequence.
 B0423651
 B0423651.1 GI:21118966
 EST.
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 870)
 NIH-MGC http://mgs.nci.nih.gov/
 Unpublished (1999)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 CONTACT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC/DCDP
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
 http://image.llnl.gov
 Plate: ILNL3342 row: 1 column: 21
 High quality sequence stop: 710.

FEATURES

source

Location/Qualifiers
 1. .870
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAG:6065828"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="N1H MGC 72"
 /note="Organ: skin; Vector: pCMV-SPORT6, Site 1: NotI, Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."

ORIGIN

Alignment Scores:

Pred. No.: 7.29e-152 Length: 870
 Score: 1447.00 Matches: 285
 Percent Similarity: 98.62% Conservative: 1
 Best Local Similarity: 98.28% Mismatches: 2
 Query Match: 68.09% Indels: 2
 DB: 5 Gaps: 0

US-10-649-273-2 (1-414) x B0423651 (1-870)

111 AAlaIeaIaThrTrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPhe 130
 Db 1 GCAATTCGAACTACCATTAACCAAGACGCTTTCCTTAAGCTGGAGTGGCTTTCATTT 60
 131 SerLeuGlnLeuValGlyGlnLeuLysProPheIleProIleHisIleMetGluAla 150
 Db 61 AGCTTACAGCTGGTGAAGACAGTTAAAGACCATTCATTCATTCATTCATTCATTCAT 120
 151 HisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIle 170
 Db 121 CATGACACTACTATTAAGGTGACCAATAAAGATTAATTCCTTTTATGTTCTTTGATT 180
 171 SerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGly 190
 Db 181 TCTGGAGGTCACTGTCTGTGGCATTAAGTCAAGAGATTCAGATTTCTGTCTTGGAA 240
 191 LysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu 210
 Db 241 AAGTTTGGACATAGCAACAGCTGACATGCTTGAACAAGTGGCAAGAAGACTTCTTTA 300
 211 IleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLys 230
 Db 301 ATAAACATCCAGAGTGTCCACCATGAGTGGGAAAGCCATAGAACATTTGGCCAAA 360
 231 GlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLysAsnGlyAsp 250
 Db 361 CAAGGAAATAGATTTTCATTTTGAACATCAAACTCCCTTCATCAGTCATTAATAATGTGAT 420
 251 PheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGlyLysGly 270
 Db 421 TTTCTTTTACCTGAGCTTCAACACCTTACTGATTAATAATGAAAAAGAAAAAGAG 480
 271 GlnGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGln 290
 Db 481 GAAGGATATTGAGAAAGGGGCAAAATCTGTCTTCAAGAGCAGACATTTGCTCCACAGTACG 540
 291 HisThrMetAlaCysHisLeuValLysArgThrHisAlaGlnIleLeuPheCysLysGln 310
 Db 541 CACCAAAATGGCATGTCATCTGTGAAAAGAACATCGGGCATTTCTGTTTGTAAAGCAG 600
 311 ArgAspLeuLeuProGlyAsnAsnAlaValLeuValAlaSerGlyValAlaAlaSerAsn 330
 Db 601 AGAGACTGTTCCTCAAAATATATGCACTGCTGTTGATCTGTGGTGTCCGACAGTAC 660
 331 PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuGly 350
 Db 661 TTCTATATCCGAGAGCTCTGGAATTTTAAACAAAGCAGACAGTGCACCTTTGGTGTGT 720

FEATURES

source

Location/Qualifiers
 1. .852
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK001YB02"
 /cell_type="HELA CELLS COR 25-NORMALIZED"
 /clone_id="HELA"
 /note="HELA cells were primed with a NotI-oligo(dT) primer. Five prime and enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
 Pred. No.: 2.41e-141 Length: 852
 Score: 1353.50 Matches: 274
 Percent Similarity: 96.14% Conservative: 0
 Best Local Similarity: 96.14% Mismatches: 8
 Query Match: 63.69% Indels: 4
 DB: 5 Gaps: 2

US-10-649-273-2 (1-414) x BX391919 (1-852)
 72 HisLeuLys-----ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsn 90
 Db 850 CATTTAACACAGGTGATGAT-----CTCCAGCAGCTCAACAGCTTCAACAGAGA-AAT 798
 91 IleglnArgIleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSer 110

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 On May 13, 2003 this sequence version replaced gi:30611736.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1240.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?cs=CS0BA10352B07_CS03317_1&c=1240.r

FEATURES

source

Location/Qualifiers
 1. .852
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK001YB02"
 /cell_type="HELA CELLS COR 25-NORMALIZED"
 /clone_id="HELA"
 /note="HELA cells were primed with a NotI-oligo(dT) primer. Five prime and enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
 Pred. No.: 2.41e-141 Length: 852
 Score: 1353.50 Matches: 274
 Percent Similarity: 96.14% Conservative: 0
 Best Local Similarity: 96.14% Mismatches: 8
 Query Match: 63.69% Indels: 4
 DB: 5 Gaps: 2

US-10-649-273-2 (1-414) x BX391919 (1-852)
 72 HisLeuLys-----ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsn 90
 Db 850 CATTTAACACAGGTGATGAT-----CTCCAGCAGCTCAACAGCTTCAACAGAGA-AAT 798
 91 IleglnArgIleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSer 110

Db 797 ATTCAACAAGATGACAGAGAGCTCTTCTGCGAGTGAAGTCTCTCCAAAGTACCTCTCA 738
 Qy 111 AAlaIeAlaThrThrIleYsPProGlyLeuAlaLeuSerIeuGlyValGlyLeuSerPhe 130
 Db 737 GCAATTGCANCTACCAATTAACCCAGNACGCTCTTTAAGCTTGAGAGTGGGCTTATCATTT 678
 Qy 131 SerIeuGlnLeuValGlyGlnLeuYsYsPProPheIleProIleHisIshMetGluAla 150
 Db 677 AGCTTACAGCTGTGTAGAGACAGTTAAABAAAGCCATTCATTCATTCATATATGAGGGCT 618
 Qy 151 HisAlaLeuThrIleArgLeuThrAsnYsValGluPheProPheLeuValLeuLeuIle 170
 Db 617 CATGACACTTACTATTAGTTGACCAATTAAGAAATTTCTTTTATGTTCTTTTGAAT 558
 Qy 171 SerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGly 190
 Db 557 TCTGGAGGTCACTGTCTGTGGCATTTAGTTCAAGGAGTTTCAGATTTCTGCTTCTTGA 498
 Qy 191 LysSerLeuAspIleAlaPProGlyAspMetLeuAspYsValAlaArgArgLeuSerLeu 210
 Db 497 AAGTTTGGACATGACACAGTGCATGCTTGCACAGTGGCAAGAGACTTTCTTTA 438
 Qy 211 IleYsHisPProGlyCysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaYs 230
 Db 437 ATAAACCAATCCAGAGTGTCTCCACCATGAGTGGGAAAGCCATAGAACATTTGGCCAAA 378
 Qy 231 GlnGlyAsnArgPheHisPheAspIleYsPProProLeuHisHisAlaYsAsnCysAsp 250
 Db 377 CAAGAAATAGATTTCAATTTTACATCAAACTCTCTGCATCATGCTAAATTTGTGAT 318
 Qy 251 PheSerPheThrGlyLeuGlnHisValThrAspYsIleIleMetYsYsGlyYsGln 270
 Db 317 TTTCTTTTACTGAGCTTCAACACGTTACTGATTAATTAATTAAGAAAAAGAAAGAG 258
 Qy 271 GlnGlyIleGlnYsGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGln 290
 Db 257 GAAGGATATTGAGAAAGGGCAATCTGTCTTACGACGACAGACATTTCTGCCACAGTACG 198
 Qy 291 HisThrMetAlaCysHisLeuValYsArgThrHisArgAlaIleLeuPheCysYsGln 310
 Db 197 CACCAAAAGGCAATGTCATCTTGAAAGAAACACATCGGGCTAATCTGTTTGTAAAGCAG 138
 Qy 311 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn 330
 Db 137 AGAGCTTGTATCTCAAAATATATAGTACGTAGTGTGATCTGGTGGTGTTCGAGATAC 78
 Qy 331 PheYrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuYs 350
 Db 77 TTTATATCCGACAGAGCTCTGGAATTTTAACAACGCAACAGTGCATTTGTTGTCT 18
 Qy 351 ProProProArgLeu 355
 Db 17 CCTCCTCCAGACTA 3
 RESULT 6
 B0961028 922 bp mRNA linear EST 21-AUG-2002
 LOCUS B0961028 8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423502
 DEFINITION AGNCOURT 8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423502
 5', mRNA sequence.
 ACCESSION B0961028
 VERSION B0961028.1 GI:22376506
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 922)
 NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@bbs-remail.nih.gov
 Tissue Procurement: DCTD/DTF/Gazdar

cDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LNC62603 row: d column: 15
 High quality sequence stop: 584.
 Location/Qualifiers
 1..922
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6423902"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_18"
 /note="Organ: lung; Vector: pOT87; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dT priming. directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using Zap-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN
 Alignment Scores:
 Pred. No.: 6,75e-133 Length: 922
 Score: 1279.00 Matches: 275
 Percent Similarity: 91.80% Conservative: 5
 Best Local Similarity: 90.16% Mismatches: 16
 Query Match: 60.19% Indels: 10
 DB: 5 Gaps: 3

US-10-649-273-2 (1-414) x B0961028 (1-922)
 Qy 80 ProProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeu 99
 Db 3 CCTCCAGCAGCTCCACAGCTTCCACAGAAATATTCACGAATATGACAAAGAGCTCT 62
 Qy 100 SerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrIleYsPProGly 119
 Db 63 TCTGCCAGTGAAGTCTCTCCAGTGAACCTTCAGCAATTCACATTAACATTAACAGGA 122
 Qy 120 LeuAlaLeuSerIeuGlyValGlyLeuSerPheSerIeuGlnLeuValGlyGlnLeuYs 139
 Db 123 CTTCCTTTAAGCTTGAGAGTGGGCTTATCATTTAGCTTAACGCTGAGAGCAATTAAA 182
 Qy 140 LysPProPheIleProIleHisIshMetGluAlaHisAlaLeuThrIleArgLeuThrAsn 159
 Db 183 AAGCCATTCATTCCTCATTCATATGAGAGGCTCATGCACTTACTATTAGTTGACCAAT 242
 Qy 160 LysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeu 179
 Db 243 AAGTGAATATTCCTTTTATGTTCTTTGATTTCTGAGGTCACTGTCTGTTGCAATTA 302
 Qy 180 ValGlnGlyValSerAspPheLeuLeuLeuGlyYsSerLeuAspIleAlaPProGlyAsp 199
 Db 303 GTTCAAGAGCTTCAATTTCTGCTTCTTGGAAAGTCTTTGACATAGCAACAGGAGAC 362
 Qy 200 MetLeuAspYsValAlaArgArgLeuSerLeuIleYsHisPProGlnCys-SerThrMet 219
 Db 363 ATGCTTGACAAAGTGGCAAGAGACTTCTTATATAAACAATCCAGAGTGCATTCACCAT 422
 Qy 219 tSerGlyGlyYsValAlaIleGlnHisLeuAlaYsGlnGlyAsnArgPheHisPheAspIle 239
 Db 423 GAGTGTGGAAAGCCATAGAACATTTGGCAAAAGAAATTAAGATTTGATTTGACAT 482
 Qy 239 eYsPProProLeuHisHisAlaYsAsnCysAspPheSerPheThrGlyLeuGlnHisVal 259
 Db 483 CAAACCTCCTTCATCATGCTTAATAATTTGATTTTCTTTTACTGAGACTTCAACAGT 542

QY 259 lThrAspIysilelleMetLysLysGluLysGluGlyileGlyLysGlyLysile 279
 Db 543 TACTGATTAATAATATAGAAAAGAAAAGAGAGATTAAGAAAAGGCAATCT 602
 QY 279 uSerSerAlaAlaAspIleAlaAlaThrValGlnIsthrMetAlaCysHileuVally 299
 Db 603 GTCTTCAGCAGACAGATTCGTGCACAGTACAGACACAAATGGCATGCTTGTGAA 662
 QY 299 sArgThrIsthrArgAlaIleLeuPheCysLysGlnIsthrAspLeuProGlnAsnAsnAl 319
 Db 663 AAGAAA-CATGGGGCTATCTGTTTGTAAAGAGAGACTTGTAACTCAAAATATATGC 721
 QY 319 aValIleuValAlaSer-GlyGlyValAlaSerAspPheTyrIleArgArgAlaLeuGln 339
 Db 722 AGTACTGGGTGATCTTGGGGGGGTGCGAGTAACTTCTATATCCGCAACTCTGGAAA 781
 QY 339 lLeuThr-AsnAlaThrGlnCysThr-LeuLeuCysProProArgLeuCysThr-A 358
 Db 782 AATTACAAAACCCAAACGGGGGACCTTGTGGGTCCCTCCCAACTATGACCTGG 841
 QY 358 sPanGlyIleMetIleAlaTrp-----AsnGlyIleGluArg--LeuArgAlaGlyL 375
 Db 842 ATTAATGGG---CATTAATGATGGCTGGGGAAGGAATTGAAAAAATACCTGTGGCT 898
 QY 375 eugly 376
 Db 899 TGGGG 903
 RESULT 7 701 bp mRNA linear EST 15-SEP-2000
 LOCUS BE740611 601595739P1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3949640 5',
 DEFINITION mRNA sequence.
 ACCESSION BE740611 GI:10154603
 VERSION BE740611.1
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 701)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: ILCM814 row: n column: 09
 High quality sequence stop: 701.
 FEATURES
 Source
 1..701
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3949640"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_9"
 /note="Organ: Ovary; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

Alignment Scores:

Pred. No.: 3,84e-121 Length: 701
 Score: 1173.00 Matches: 229
 Percent Similarity: 99.15% Conservative: 3
 Best Local Similarity: 97.86% Mismatches: 0
 Query Match: 55.20% Indels: 2
 DB: 2 Gaps: 0

US-10-649-273-2 (1-414) x BE740611 (1-701)

QY 125 GlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysProPhePro 144
 Db 2 GGAGTGGGCTTATCATTTAGCTTACAGCTGTGAGACAGTTAAAAAGCATTCATCCC 61
 QY 145 lIleHileMetGlnAlaIsthrIleArgLeuThrAsnLysValGlnPhePro 164
 Db 62 ATTCATCAATATGAGGCTCATGACCTTATTAAGGTACCAATAAAGAGATTTCT 121
 QY 165 PheLeuValLeuLeuIleSerGlyLysCysLeuLeuAlaLeuValGlnGlyValSer 184
 Db 122 TTTTATGTTCTTTGATTTCTGAGGTCACTGCTGTGGCATTAAGTCAAGAGATTCA 181
 QY 185 AspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspPheLeuAspLysVal 204
 Db 182 GATTTCTGCTCTTCTGAAAGTCTTTGACATATGACACAGTGCATGCTTGAACAAGTGC 241
 QY 205 AlaArgArgLeuSerLeuIleLysHileProGluCysSerThrMetSerGlyGlyValAla 224
 Db 242 GCAGAGAGACTTCTTAAATTAACATCCAGAGTGTCCACCAATGAGTGGGAAAGCC 301
 QY 225 lIleGlnHileuAlaLysGlnGlyAsnArgPheHilePheAspIleLysProPheLys 244
 Db 302 ATAGAACATTTGGCCAAACAGGAATATGATTTCAATTTGACATCAACCTCCCTTCAT 361
 QY 245 HileAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHileValThrAspLysile 264
 Db 362 CATGCTAAAAATGTGATTTTCTTTTACCTGACCTTCAACACGTTACTGTAATAATATA 421
 QY 265 MetLys--LysGluLysGluGluGlyIleGlnLysGlyGlnIleLeuSerSerAlaAla 284
 Db 422 ATGAAAAACAGAAACAGAGAGAGATTTGAGAAAGGGCAAACTCTGTTCCAGCAGCAG 481
 QY 284 sPheIleAlaAlaThrValGlnHileThrMetAlaCysHileValLysArgThrHileArg 304
 Db 482 ACATTCCTGCACAGTACAGACACATGCAATGATCTTGTGAAAAGAACATATGGG 541
 QY 304 lAlleuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlas 324
 Db 542 CTATTTCTGTTTCTAAGCAGAGACTTGTATCTCAAAATATGACATGCTGGTGCAT 601
 QY 324 eGlyGlyValAlaSerAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAla 344
 Db 602 CTGGTGGTGGCGCAATTAATCTTATATCCGACAGCTCGGAATTTTAAACAAACGCAA 661
 QY 344 hGlnCysThrLeuLeuCysProProArgLeuCys 356
 Db 662 CACAGTCACTTGTGTGTCTCTCCACAGCTATATGC 699
 RESULT 8 749 bp mRNA linear EST 16-MAR-2004
 LOCUS CK982692 4115331 BANC 9BOV Bos taurus cDNA clone 9BOV45_M02 5', mRNA
 DEFINITION sequence.
 ACCESSION CK982692
 VERSION CK982692.1 GI:45500672
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 REFERENCE 1 (bases 1 to 749)
 AUTHORS Sonstegard, T.S., Van Tassell, C.P., Matukumalli, L.K., Harhay,

TITLE
G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C.
Production of EST from cDNA libraries derived from immunologically
activated bovine gut
JOURNAL
Unpublished (2004)
COMMENT
Contact: Tad S. Sonstegard
Bovine Functional Genomics Laboratory
Animal and Natural Resources Institute
Bldg. 200 Rm2A BARC-Baet, Beltsville, MD 20705, USA
Tel: 3015048416
Fax: 3015048414
Email: tads@nri.barc.ueda.gov

Single pass sequencing. Bases called and trimmed with phred
0.00925 using options "-trim alt" -trim fasta. Vector identified
by cross match using options -mismatch 12 -mismatch 18
Plate: 45 row: M column: 02
Seq primer: CCCAGTCAAGAGCTGTAAACG
High quality sequence stop: 749.

FEATURES
Source

1. 749
Location/Qualifiers
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="9BOV45_M02"
/sex="Male"
/tissue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_1lb="BARC 9BOV"
/note="Organ: Abomasum; Vector: phgen-1; Site 1: EcorV;
Site 2: NotI; Equimolar amounts of mRNA extracted from
fundic and pyloric abomasum of 18 and 21 week old steers.
Exposure to Osterriaga osterriagi was initiated at 15 weeks
of age. fundic and pyloric abomasum"

ORIGIN

Alignment Scores:

Pred. No.: 2,096-119 Length: 749
Score: 1158.00 Matches: 225
Percent Similarity: 95.18% Conservative: 12
Best Local Similarity: 90.36% Mismatch: 12
Query Match: 54.49% Indels: 0
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CK982692 (1-749)

QY 49 ThrAlaAlaValAlaValAspGluThrGlyAsnValLeuGlyGluAlaIleHisSerGln 68
DB 1 ACAGCAGCTGCTGAGTGAAGCTGGAATGTTTGGAGAGCAATACATCCCA 60
QY 69 ThrGluValHisLeuValThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArg 88
DB 61 ACTGAGCTTCATTTAAACAGGTGGGATTTCTCCAGTAGCTCAACAGCTTATAGA 120
QY 89 GluAsnIleGlnArgGlyIleValGlnGluAlaLeuSerAlaSerGlyValSerProSerArg 108
DB 121 GAAATATTTCAACGATAGTACAGAAAGCTCTCTGTCAGAGAGTCTCTCCAGTAGA 180
QY 109 LeuSerAlaIleAlaThrThrIleHisProGlyLeuAlaLeuSerLeuGlyValGlyLeu 128
DB 181 CTCACAGAGTTCACACACATTAAGCCAGAGCTTTCATTAAGCTTGGCCTTAGTTTA 240
QY 129 SerPheSerLeuGlnLeuValGlyGlnLeuValSerProPheIleProIleHisIleMet 148
DB 241 TCGTTTAGCTTACACATGCTAGACCAAGTTTAAAGCCCTTCATTCACATTCATCAGATG 300
QY 149 GluAlaHisAlaLeuThrIleArgLeuThrAsnValValGluPheProPheLeuValLeu 168
DB 301 GAGGCTCATGACCTCATTTAGGTATTAACAATAGAGTAGAATTTCCGTTTATGTTCTT 360
QY 169 LeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeu 188
DB 361 TTGATTTCTGAGGCTATGCTTTTGGCATTTAGTAGAGGAGTTTCAAGATTTTCTCTT 420

QY 189 LeuGlyLeuSerLeuAspIleAlaProGlyAspMetLeuAspValAlaArgArgLeu 208
DB 421 CTGGGAAGTCTTTGGACATAGCACAGGAGCATGCTTACACAGTAGAAGAAAGACTT 480
QY 209 SerLeuIleValHisPheProGluCysSerThrMetSerGlyGlyValAlaIleGluHisLeu 228
DB 481 TCTTATATAAATCATCCAGAGTCTCCACCATGATGAGCGGAGAGCTATAGAACTTTTG 540
QY 229 AlaValGlnGlyAsnArgPheHisPheAspIleLeuProProLeuHisHisAlaValAsn 248
DB 541 GCCAAACAGGAAATGATGATTCATTTGATTTCCAGCTCCCATGCAACGCTTAAAAAT 600
QY 249 CysAspPheSerPheThrGlyLeuGlnHisValThrAspValIleIleMetValGlyGlu 268
DB 601 TGTGATTTTCTTTTCTGACCTTCAACGTTATGATGATGATGATGATGATGATGATGATG 660
QY 269 LysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaThr 288
DB 661 AAAGAGGAAGTATGAGACAGGCGGAGGCTCTGCTTCAAGCTGCAACATTCGTCGCG 720
QY 289 ValGlnHisThrMetAlaCysHisLeu 297
DB 721 GTGACAGCACCGTGCCTGCCACATTT 747

RESULT 9
COJ38006

LOCUS COJ38006 800 bp mRNA linear EST 29-JUL-2004
DEFINITION SLH03c19e02f1 squirrel heart library 1 Sperophilus lateralis cDNA
clone 19e02 5', mRNA sequence.

ACCESSION COJ38006
VERSION COJ38006.1 GI:50825276

KEYWORDS EST.
SOURCE Sperophilus lateralis (golden-mantled ground squirrel)

ORGANISM

Sperophilus lateralis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scuriinae;
Sperophilus.

1 (bases 1 to 800)

Williams,D.R., Gracey,A.Y., Martin,S.L., Hughes,M.A., Li,W.,
Rogers,J. and Cousins,A.R.

Microarray analysis of transcriptional changes during hibernation
in the golden mantled ground squirrel, Sperophilus lateralis

Unpublished (2004)

Contact: Andrew R. Cousins
Laboratory for Environmental Gene Regulation

University of Liverpool
School of Biological Sciences, The Biosciences Building, Crown
Street, Liverpool, United Kingdom, L69 7GB

Tel: +44(0)151-795-4510

Fax: +44(0)151-795-4431

Email: cousins@lly.ac.uk
Vector has been trimmed from this EST.

Plate: 19 row: e column: 02
Seq primer: pfic T7 (5'-ATACGACTCACTATAGG-3')

High quality sequence stop: 800.

Location/Qualifiers

FEATURES

source

1. 800
/organism="Sperophilus lateralis"

/mol_type="mRNA"

/db_xref="taxon:76772"

/clone="19e02"

/sex="Male & female"

/tissue_type="Heart"

/dev_stage="Adult"

/lab_host="R. coli Electromax DH10B"

/clone_1lb="squirrel heart library 1"

/note="Vector: pFUC; Site 1: Salt GCGAG; Site 2: BamHI
GATTC; NotI; and subcloned cDNA library prepared
from heart of hibernating and summer animals"

ORIGIN

Alignment Scores:

Pred. No.: 8.74e-118 Length: 800
 Score: 1144.00 Matches: 226
 Percent Similarity: 97.05% Conservative: 4
 Best Local Similarity: 95.36% Mismatches: 7
 Query Match: 53.84% Indels: 0
 DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CO738006 (1-800)

QY 1 MetleuileleuThrlySerThraAlaGlyValPhePheIysProSerIysArgIysValTyr 20
 Db 88 ATGCTAATATTGAATTAAGACACCGAGATTCTTTTAAACCAATCCAAAGAAAGATTGAAT 147
 QY 21 GluPheLeuAgsSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValIleVal 40
 Db 148 GGATTTTAAAGAGTTTAAATTTTCACTCGAACAACATTTTCTCATTAATAATAGTATTG 207
 QY 41 GlyIleGlyIleThrSerCysAspAspThrAlaAlaAlaValAlaAspGlyThrGlyVal 60
 Db 208 GGTATTGAAACAGCTGTGATGATACAGACGCTGGGTGTGATGAATGAATGGAATGTG 267
 QY 61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisIleuIysThrGlyGlyIleValPro 80
 Db 268 TTGGGAGAGCAATACATTTCCCAACTGAAGTTCAATTAAACAGGTGGGATATTCTCT 327
 QY 81 ProAlaIleGlnGlnLeuHisArgGlyIleValIleValIleValIleValIleValIle 100
 Db 328 CCAGTAGCTCAACATTCATTAAGAAACATTCACAGAAATGTAACAAAGAGCTTTTGT 387
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
 Db 388 GCCAGTGAATATCTCCAAAGTGAATCTCAGCGATTGCAACATTAAGCCAGAGACTT 447
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIys 140
 Db 448 GCTTTAAGCTTAGAGGTAGGTGCTTATCATTTAGTTACAGCTGTGTAGACACGCTGAAAAG 507
 QY 141 ProPheIleProIleHisIleMetGlyAlaHisAlaLeuThrIleArgLeuThrAsn 160
 Db 508 CCATTTCATCCCATTCATCATTAAGAGGCTCAGCACTTACTATTAGGTGACCAATATA 567
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuValIleVal 180
 Db 568 GTAGAGTTTCCATTTTATGTTCTTTGATTTCTGAGGTCACTGCTCTTGSCATTAGTT 627
 QY 181 GlnGlyValSerAspPheLeuLeuGlnGlyIysSerLeuAspIleAlaProGlyAspPhe 200
 Db 628 CAAGGAGTTTCAGATTTTCTGCTCTTGGAAGTCTTTGACATTCACACGAGTCAATG 687
 QY 201 LeuAspIysValAlaArgArgLeuSerLeuIleIysHisProGlyIysSerThrMetSer 220
 Db 688 CTTACCAAGGTAGCAAGAAAGCTTTTAAACCAACATCCAGAAATGCTCCACATGAGT 747
 QY 221 GlyGlyValAlaIleGlnHisLeuAlaIysGlnGlyAsnArgPheHisPhe 237
 Db 748 GCGCGGAAGGCTATTAAGCAATTTGGCCAAACAGAAACAGATTTTCATTTT 798

RESULT 10
 BM907988 1109 bp mRNA linear EST 12-MAR-2002
 LOCUS BM907988
 DEFINITION AGENCOURT_6707465 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5744934
 5', mRNA sequence.
 ACCESSION BM907988
 VERSION BM907988.1 GI:19358367
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1109)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed By: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing By: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM12767 row: b column: 07
 High quality sequence stop: 602.
 Location/Qualifiers
 1..1109

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5744934"
 /tissue_type="medulla"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 119"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: EcoRV (destroyed); RNA source normal medulla from
 anonymous male age 27. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.3 kb, insert size range
 0.9-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 013. Note:
 this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 3.28e-116 Length: 1109
 Score: 1132.00 Matches: 241
 Percent Similarity: 88.89% Conservative: 15
 Best Local Similarity: 83.68% Mismatches: 23
 Query Match: 53.27% Indels: 9
 DB: 5 Gaps: 4

US-10-649-273-2 (1-414) x BM907988 (1-1109)

QY 1 MetleuileleuThrlySerThraAlaGlyValPhePheIysProSerIysArgIysValTyr 20
 Db 129 ATGCTAATATTGAATTAAGACACCGAGATTCTTTTAAACCAATCCAAAGAAAGATTGA 188
 QY 21 GluPheLeuAgsSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValIleVal 40
 Db 189 GAATTTTAAAGAGTTTAAATTTTCACTCGAACACTAATTTCTCATTAATAATAGTATTG 248
 QY 41 GlyIleGlnThrSerCysAspAspThrAlaAlaAlaValAlaAspGlyThrGlyVal 60
 Db 249 GGAAATTTGAAAGAGTTGTATATATACAGACGCTGTGTGTATGAATGGAATGAGTGTG 308
 QY 61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisIleuIysThrGlyGlyIleValPro 80
 Db 309 TTGGGAGAGCAATACATTTCCAAACTGAAGTTCAATTTAAACACAGTGGAGATTGTTCT 368
 QY 81 ProAlaIleGlnGlnLeuHisArgGlyIleValIleValIleValIleValIleValIle 100
 Db 369 CCAGGAGCTCAACAGATTCAAGAGAAATATTCACGAATAGTACAGAAAGCTTTTCT 428
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
 Db 429 GCCAGTGGAGTCTCCCAAGTACCTCTCAGCAATTTGCAACTACATTAACCAAGACCTT 488
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIys 140
 Db 489 GCTTTAAGCTTAGAGGTAGGTGCTTATCATTTAGCTTACAGCTGTGTAGACAGTTAAAG 548
 QY 141 ProPheIleProIleHisIleMetGlyAlaHisAlaLeuThrIleArgLeuThrAsn 160
 Db 549 CCATTTCATCCCATTCATCATTAAGAGGCTCAGCACTTACTATTAGTTGACCAATATA 608

QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 DB 609 GTAGAAATTCCTTTTGTAGTCTTTGATTCTGGAGGTCACTGCTGTGGCATAGTT 668
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 DB 669 CAAGAGATTGATGATTTCTGCTCTGAAAGGCTTGGACATTAACACAGGGGCAATG 728
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer 220
 DB 729 CTGTAACAAGGGGCAAGAAAGATTCTTTATTAACAATCCAGAGGCTCCACCATGAG 788
 QY 221 GlyGlyLysAlaIleGluHisLeu--AlaLysGlnGlyAsnArgPheHisAsp--Asp 238
 DB 789 GGGGGGAAACCATTAACCATTTGGGCAACCAAGAAATAGAATTTCTTTTGGAAAT 848
 QY 239 IleLysProPheLeuHisHis-AlaLysAsnCysAspPheSerPheThrGly--LeuGln 257
 DB 849 CCAAACTCCCTGATTAATGTTAAATAAGGATGGAATTTTCTTAAATGGGAATTTCAA 908
 QY 258 HisVal--ThrAspLysIleIleMetLysGlyLysGluGluGlyIleGlu----- 274
 DB 909 CACGCTTAAGTGGTTAAATATATGAGAAAGGAAAGAGAGGATTTGGGA 968
 QY 275 --LysGlyGlnIleLeu 279
 DB 969 AAAGGGGCAAAATCTCTG 986

RESULT 11
 LOCUS CF114247 765 bp mRNA linear EST 23-JUL-2003
 DEFINITION Shultzomc07498 Rat lung airway and parenchyma cDNA libraries
 VERSION CF114247 Rat mus norvegicus cDNA clone NA4935 5', mRNA sequence.
 KEYWORDS CF114247.1 GI:33173962
 SOURCE EST.
 ORGANISM Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 765)
 Shultz,M.A., Zhang,L., Gu,Y.-Z., Baker,G.L., Fannuchi,M.V.,
 Padua,A.M., Gurske,W.A., Morin,D., Penn,S.G., Jovanovich,S.B.,
 Plopper,C.G. and Buckpitt,A.R.
 Gene expression analysis in response to lung toxicants: I.
 Sequencing and microarray development
 Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)
 Contact: Shultz MA
 Dept. of Molecular Biosciences, School of Veterinary Medicine
 University of California, Davis
 111 Haring Hall, One Shields Avenue, Davis, CA 95616, USA
 Tel: 530 752 0793
 Fax: 530 752 4698
 Email: mashultz@ucdavis.edu

Average Phred score is 20 or better. All poor quality data (Phred <
 20) and vector/linker sequence has been removed.
 High quality sequence stop: 765.
 Location/Qualifiers

FEATURES
 source 1..765
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="NA4935"
 /sex="male"
 /tissue_type="airway or parenchyma"
 /dev_stage="adult"
 /clone_id="Rat lung airway and parenchyma cDNA libraries"
 /note="Organ: lung; Vector: pGEM-11zf(-); Site 1: Eco RI;
 Site 2: Not I; mRNA was isolated from microdissected rat
 lung airways and parenchyma tissues."

ORIGIN

Alignment Scores:
 Prod. No.: 4,26e-114 Length: 765
 Score: 1111.00 Matches: 216
 Percent Similarity: 93.78% Conservative: 10
 Best Local Similarity: 89.63% Mismatches: 15
 Query Match: 52.28% Gaps: 0
 DB: 7

US-10-649-273-2 (1-814) x CF114247 (1-765)

QY 24 ArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValIleGlu 43
 DB 39 AAGAAGATTATGTTCAACTAGACGACCTTTTCATCATTAACCTGGGCAATTGAA 98
 QY 44 ThrSerCysAspAspPheThrAlaAlaValAlaLysGluThrGlyAsnValIleGlu 63
 DB 99 ACCAGCTGTGATACACAGACGCTGCTGTGGATGAAGAACTGGAAATGCTGGAGAA 158
 QY 64 AlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValProProAla 83
 DB 159 GCACTCACTCCAGACGAGAGTCAATCGAAACAGTGGAGTGTCTCTCAAGTCGCT 218
 QY 84 GlnGlnLeuHisArgGluAsnIleGlnArgIleValIleGlnIleValLeuSerAlaSerGly 103
 DB 219 CAACAGCTTCACAGAGAAATATTCAGCGAATGTAGAGAAAGCTCTTCTGCCAGTGG 278
 QY 104 ValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSer 123
 DB 279 GTCTCCCAAGGACCACTCTCAGCAATTTGCAATCAATCAACCAAGACTGGCCCTAAGC 338
 QY 124 LeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLysProPheIle 143
 DB 339 CTGGGGGTGGCTTATCTTTAGCGTACAGTAAATCAATTAAGCAATTTATC 398
 QY 144 ProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPhe 163
 DB 399 CCATTCATCAATGAGAGCTCAGCGCTGACTATTAGGTGACCAACAAAGTCGATTT 458
 QY 164 ProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnIleVal 183
 DB 459 CTTTATTGATCTTTTGAATTTCTGGAGGCCACTGCTGTGGCGTTAGTTCAGAGTGT 518
 QY 184 SerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyLysPheLeuAspLys 203
 DB 519 TCAGATTTTCTCTCTCTCGGAAAGTCCCTGGAATGACCCAGGCCACATGCTTGA 578
 QY 204 ValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSerGlyGlyLys 223
 DB 579 GTGGCAAGAGACTTTCTTTATCAAAATCCAGAAATGTTCTTACATGAGATGGGAA 618
 QY 224 AlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProPheLeu 243
 DB 639 GCTATAGAAATTTGGCCAAAGAGAAATAGATTTCACTTACTATCATATCACCACATG 698
 QY 244 HisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIle 263
 DB 699 CAGAATGTGAAGAACTGTATTTTCTTTTACGGGACTTCAACATGTCAACCGATAGCTA 758
 QY 264 Ile 264
 DB 759 ATA 761

RESULT 12
 LOCUS B0636028 640 bp mRNA linear EST 15-JUL-2002
 DEFINITION h03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
 B0636028
 VERSION B0636028.1 GI:21760487
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PIRMBED
COMMENT
FEATURES
SOURCE

1 (bases 1 to 640)
Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A.,
Touchman, J.W., Bouffard, G., Smith, D., and Peterson, K.
Expressed sequence tag analysis of human retina for the NEIBank
Project: Retbindn, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts
Mol. Vis. 8 (4), 196-204 (2002)
22103461
12107411

Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 03 row: d, column: 11
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers

1..640
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hd03d11"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retina cDNA (Un-normalized,
unamplified): hd/he"
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
was dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was constructed at
Bioserve Biotechnology (Laurel MD) essentially following
details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor
[5'-pGACTGAGTCTGACGAGCGGCGCC(T)15-3']. EST analysis
was performed on the unamplified library at the NIH
Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
Pred. No.: 1..228-112 Length: 640
Score: 1097.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 51.62% Indels: 0
DB: 5 Gaps: 0

US-10-649-273-2 (1-414) x B0636028 (1-640)

QY 206 ArgArgLeuSerLeuIleIyShiSProGluCySserThmetSerGlyIyLysAlaIle 225
Db 3 AGAAGACTTTCTTTATATAAAACATCCAGAGTCCTCCACATGAGTGTGGAAAGCCATA 62
QY 226 GluHisLeuAlaIyGluGluGlyAsnArgPheHisSphEAspIleIyEProProLeuHisShis 245
Db 63 GAACATTTGGCCAAACAGAGAAATAGATTTCATTTGACATCAAACTCCCTTGACATCAT 122
QY 246 AlaIyAsnCyAspPheSerPheThrgIyLeuGluHisValThrAspIyIleIleMet 265
Db 123 GCTAAAAATTGATTTTCTTTTACTGACATTCACACGTTACTGATATAAATTAATG 182
QY 266 LysIyGluIyGluGluGlyIleGluIyGluGluIleLeuSerSerAlaIaAspIle 285
Db 183 AAAAAGAAAAAGAGAGATTTAGAAAGGGCAATCTCTGCTTCAACAGACATTT 242

QY 286 AlaAlaThrValGluHisThrMetAlaCyShiSLeuValIyEArgThrHisArgAlaIle 305
Db 243 GCTGCACAGTACAGACACATGGCATCTTGTGAAAAAGACACATGGGCTATT 302
QY 306 LeupheCyAlyGluArgAspLeuLeuProGluAsnAsnAlaValLeuValAlaSerGly 325
Db 303 CCGTTTGTAGACAGAGACATCTTCACTCAAAAATAGCATCTGGTGCATCTGGT 362
QY 326 GluValAlaSerAspMetIyIleArgArgAlaIleGluIleLeuThrAsnAlaThrGln 345
Db 363 GGATGCAAGTACTTCTATATCCGACAGCTTGAAATTTTAAACAGACACAG 422
QY 346 CyThrLeuLeuCySProProArgLeuCySthrAspAsnGlyIleMetIleAlaArg 365
Db 423 TCACCTTTTGTGTCTCTCTCCACATGACATGATTAAGGATTTATGATTCATGG 482
QY 366 AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArg 385
Db 483 AATGGTATGAAAGACATGACGTGCTGGCATTTTACATGACATGAGGACATCCGC 542
QY 386 TyrGluProIyCySProLeuGlyValAspIleSerIyGluValGlyIleAlaSerIle 405
Db 543 TATGAACCAAAATGTCCTCTTGAGTACATATCAAAAGAGTGTGAGAGCTTCATA 602
QY 406 LysValProGluLeuIySmetGluIle 414
Db 603 AAAATACCAATTAATAATGACAGATA 629

RESULT 13
CN823245 730 bp mRNA linear EST 02-JUN-2004
LOCUS CN823245
DEFINITION Oa splhn 04N08 M13reverse Sheep spleen\brain pSPORT1 library Ovis
arles cDNA clone Oa_splhn_04N08 5', mRNA sequence.
ACCESSION CN823245
VERSION CN823245.1 GI:47951314
KEYWORDS EST.
SOURCE Ovis aries (sheep)
ORGANISM Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Caprinae; Ovis.
1 (bases 1 to 730)
Gossner, A. and Hopkins, J.
Ovine spleen\brain cDNA library
Unpublished (2004)
CONTACT: J Hopkins
Veterinary Biomedical Sciences
University of Edinburgh
Summerhall Square, Edinburgh, EH9 1QH.
Email: j.hopkins@ed.ac.uk
Plate: 04 row: N column: 08
Seq primer: M13reverse
High quality sequence start: 6
High quality sequence stop: 550.
Location/Qualifiers

FEATURES
source 1..730
/organism="Ovis aries"
/mol_type="mRNA"
/db_xref="taxon:9940"
/clone="Oa_splhn_04N08"
/clone_lib="Sheep spleen\brain pSPORT1 library"
/note="Vector: pSPORT1"

ORIGIN

Alignment Scores:
Pred. No.: 1..088-106 Length: 730
Score: 1045.00 Matches: 204
Percent Similarity: 93.51% Conservative: 12
Best Local Similarity: 88.31% Mismatches: 15
Query Match: 49.18% Indels: 0
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CN823245 (1-730)

```

QY 66 GlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThr 115
DB 37 CAAGAAGCTCTCTGCGCCAGTAAAGTCTCTCCAGAGAACTCTCAGCAATTCGAACTTACC 96
QY 116 IleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuVal 135
DB 97 GTAAAGCCAGACCTTCTTAAAGCTTGGCGGTAGGTTTATCATTTAGCTTACACCTGGTA 156
QY 136 GlyGlnLeuLysLysProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIle 155
DB 157 GACCAAGTTTAAAGCCCTTCATCCATTCATCATGAGAGCTCATGACCTACTACTATT 216
QY 156 ArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCys 175
DB 217 AGGTTAAACAAATTAAGTAAAGTAAATTTCCATTTTATTTGATTTCTGAGGTCATTTGT 276
QY 176 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 195
DB 277 CTTTGGCATTAGTACAGAGAGATTTCAGATTTTCTTCTTGGAAAGCTTTGGACATA 336
QY 196 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 215
DB 337 GCACACAGCGCAGATGCTTGACAGAGTAGCAGAGACCTTGTATATAAATCAATCCAGAG 396
QY 216 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 235
DB 397 TGTCTCCACCAAGAGGCGGAGGCTATAGAACTTTGGCCAAACAGGAAATAGATTG 456
QY 236 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 255
DB 457 CATTTCATTTTCCAACTCCATCCAGTACAGTGTCTAAATAATTTGATTTTCTTTCTGGA 516
QY 256 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyIleGlyLys 275
DB 517 CTTCAACATGTTATGATTAAGTATGATTAATGCAAGAGAAAGAGAGAGATTCAGAGCAG 576
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295
DB 577 GGGCAGATCCCTGCTTTCAGCTGACAGATGCTGCTGCAGTCCACACACCTTACGCTGC 636
QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315
DB 637 CACGTTCGAAAGAAACACATGCGCTATCTCTGTCGACAGACAGAGCGCTTGTACT 696
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGly 326
DB 697 CAAGTAACGACGACTGTTGTATCTGGAGGC 729

RESULT 14
CK833139 661 bp mRNA linear EST 04-MAR-2004
LOCUS 4057170 BARC 8BOV Bos taurus cDNA clone 8BOV_20020 5', mRNA
DEFINITION
sequence.
ACCESSION CK833139 GI:45061881
VERSION CK833139.1
KEYWORDS Bos taurus (cow)
SOURCE EST.
ORGANISM Bos taurus
Bos taurus
Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 661)
Baumann,R.G., Baldwin,R.L., Sonsegard,T.S., Van Tassel,C.P. and
Matukumalli,L.K.
Construction and Analysis of a cDNA Library Generated From
Intestinal Muscle and Epithelial Tissues of Holstein Cattle
Unpublished (2004)
Contact: Richard G. Baumann
Bovine Functional Genomics Lab
ANRI
BUDG 162: BARC-EAST, Beltsville, MD 20705, USA
Tel: 3015048604

```

```

FEATURES
    source
        1..661
            /organism="Bos taurus"
            /mol_type="mRNA"
            /strain="Holstein"
            /db_xref="taxon:9913"
            /clone="8BOV_20020"
            /sex="female"
            /tissue_type="Epithelial, Muscle"
            /dev_stage="lactating, Neonatal"
            /lab_host="DH10B Tona"
            /clone_1lb="BARC 8BOV"
            /note="Organ: Intestine; Vector: pCMVSPORT6.1, Site_1:
            NotI; Site_2: EcoRI; Normalized cow cDNA intestinal
            library in pCMVSPORT6.1, constructed from equimolar mRNA
            pools derived from 5 sources, 4 lactating intestinal, 1
            neonatal intestinal 4/5 lactating, proximal duodenum,
            jejunum, distal ileum, colon, 1/5 neonatal, proximal
            duodenum, jejunum, distal ileum"

ORIGIN
Alignment Scores:
Pred. No.: 4,98E-106 Length: 661
Score: 1038.50 Matches: 203
Percent Similarity: 95.48% Conservative: 8
Best Local Similarity: 91.86% Mismatches: 9
Query Match: 48.87% Indels: 1
DB: 7 Gaps: 1

US-10-649-273-2 (1-414) x CK833139 (1-661)
QY 30 ProGlyThrLeuPheLeuHisLysIleValIleGlyIleGluThrSerCysAspAspThr 49
DB 2 CCGGG--ATATTCCTTCTATAACTAGTATTGGGAATGAAACGCTTGATGATGACA 58
QY 50 AlaAlaAlaValValAspGluThrGlyAsnValLeuGlyGlnAlaIleHisSerGlnThr 69
DB 59 GAGGCTGCTGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 118
QY 70 GluValHisLeuLysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGln 89
DB 119 GAAGTTCATTATAAAGCAGGTGGATTAATTCCTCCAGTACCTCAACGCTTCATAGGAA 178
QY 90 AsnIleGlnArgIleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeu 109
DB 179 AATATTCAACGCACTAGTACAGAAAGCTCTCTGCGCAGTAAAGTCTCTCAAGTGAATC 238
QY 110 SerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSer 129
DB 239 TCAGCAGTTGCAACCACTAAAGCCAGACCTTGAAGCTTGGCGGTAGGTTATCG 298
QY 130 PheSerLeuGlnLeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetGln 149
DB 299 TTTAGCTTCAACCTGTAGACCAAGTTTAAAGCCCTTCATTCCTCAATTCACATGAG 358
QY 150 AlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeu 169
DB 359 GCTCATGCACTCACTATAGGTTAAACAAATTAAGTAAATTTGATTTGATTTG 418
QY 170 IleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeu 189
DB 419 ATTTCGAGAGCTATGCTTTTGGCATTAGTAAAGAGATTTCAGATTTCTTCTTCTT 478
QY 190 GlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSer 209

```

Fax: 3015048744
 Email: rbaumann@anri.barc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 0.000925 using options -trim_alc - -trim_faaba. Vector identified
 by cross match using options -minmatch 12 -minscore 18
 plate: 20 row: 0 column: 20
 Seq primer: CTTATTAGGTGACACTATGAAAC
 High quality sequence stop: 661.

Db 479 GGGAAGCTTTGGACATAGACACAGGTGACATGCTTGAACAAGTAGCAAGACTTCC 538
Qy 210 Leuilelyshisproglucysethmetserglylyvalallegluhiislemja 229
Db 539 TTAATTAATAACCAAGAGTGTCTCCACCATGATGCGGAGGCTATAGAACATTTGGCC 598
Qy 230 Lyseglnglyaenarphneishisphesapilleylspropoleuhihisalalyasancys 249
Db 599 AAACAGGAATATAGATTGCATTTCATTTCACGCTCCCATGCAACGTCCTAATAATGT 658
Qy 250 Asp 250
Db 659 GAT 661

RESULT 15
BQ433135 879 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT 7760756 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6016098
DEFINITION 5', mRNA sequence.
ACCESSION BQ433135
VERSION BQ433135.1 GI:21172211
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgsb@nsl.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LHM13213 row: d column: 19
High quality sequence stop: 674.
Location/Qualifiers
1..879
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6016098"
/issue_type="embryonal carcinoma, cell line"
/lab_host="RDH108 (phage-resistant)"
/clone_lib="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

ORIGIN
Alignment Scores:
Pred. No.: 1.39e-103 Length: 879
Score: 1018.50 Matches: 210
Percent Similarity: 94.67% Conservative: 3
Best Local Similarity: 93.33% Mismatches: 9
Query Match: 47.93% Indels: 3
DB: 5 Gaps: 1

US-10-649-273-2 (1-414) x BQ433135 (1-879)

Qy 1 Metleuileuthrlysthralaglyvalpnephelysproserilyearglyvaltyr 20
Db 201 ATGCTAAATCTGACTAAGACTGCAAGAGTTTTTTTAAACCATCAAAAGAAAGTTTAT 260
Qy 21 Glupheleuargserphneishisphesapilleylspropoleuhihisalalyasancys 40

Db 261 GAATTTTAAAGATTGTTAATTTCATCTCGAACAATAATTTCTTCATAAATAATGATTG 320
Qy 41 GYIIEgluThSerCyasphspphralaalavalvalaspqlurhrglyaanval 60
Db 321 GGAATTTGAAATAGTTGTATATATACAGACGCTGCTGTGTGATGAACTGGAAATGTG 380
Qy 61 Leuglyglualalishisserglnthrglyvalhlsleuylserhrglylylevalpro 80
Db 381 TTGGAGAAAGCAATATCATTTCCCAACTGAAGTTCAATTTAAACACAGTGGATGTTCCT 440
Qy 81 Proalalaglnglnleuhi salrglyuanll eglnarglleyalnglnualaleuser 100
Db 441 CCAGCAGCTCAACAGCTTCACAGAAATATTCACGAATAGTACAGAAAGCTCTTCT 500
Qy 101 AlaSerglyvalserproserapleuseralalialaThrThrlyleyproglyleu 120
Db 501 GCCAGAGAGTCTCTCCAAAGTACCTCTGACGAATTCACATCAATTAACAGACTT 560
Qy 121 AlaLeuserLeuglyvalglyleuserpheserLeuGlnleuValglyGlnleuylslys 140
Db 561 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACACTGGTAGACAGCTTAAAAAAG 620
Qy 141 Propheileproilehiswecglualahlsalaleuthrleargleuthraanlys 160
Db 621 CATTCATTCATTCATCATATGAGAGCTCATGCACTTACTATAGTTGACCAATTA 680
Qy 161 ValGluPherProphleuValleuuleiserglyglyhihsyaleuValaleuVal 180
Db 681 GTAGAAATTCCTTTTATGTTCTTTGATTCGTGAGGTCACATGTCGTGGCATTAAGT 740
Qy 181 GlnGlyValSerAspPhleuLeuLeuGlylylserLeuAspillealaprogllyaspmet 200
Db 741 CAAGAGTTTCAGATTTTCTGCTCTTGAAAGCTCTTGACATACACAGATGACATG 800
Qy 201 -LeuaspLyValAlaArgArgLeuserLeuLeu--lyshis-ProGluCyserThrm 219
Db 801 GCTTGCAAGGTGGGCAAAAGACTTCTTATATANAACATTCACAAAGTCTCCCCC 860
Qy 219 etserglygly 222
Db 861 ATGGAAMGGGT 871

Search completed: February 16, 2005, 21:04:35
Job time : 4943.95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 10:46:17 ; Search time 70.2 Seconds
(without alignments)
3019.954 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 2125
Sequence: 1 MLILTKTAGVFPKSKRKRY.....DISKEVGASIKVPLQKMEI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_tr embl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	2125	100.0	096EV9	096ev9 homo sapien
2	2090.5	98.4	09H4B0	09h4b0 homo sapien
3	1845	86.8	096NH5	096nh5 homo sapien
4	1835	86.4	09PEB4	09peb4 mus musculu
5	1827	86.0	09BLB6	09blb6 mus musculu
6	1819	85.6	09DON0	09don0 mus musculu
7	1725	81.2	06AYN7	06ayn7 rattus norv
8	1283	60.4	08JFW3	08jfw3 brachydanio
9	1279	60.2	08JFR7	08jfr7 brachydanio
10	772.5	36.4	09VWD6	09vwd6 drosophila
11	714.5	33.6	09VWD6	09vwd6 drosophila
12	668.5	31.5	022145	022145 drosophila
13	577.5	27.2	073H71	073h71 wolbachia p
14	568.5	26.8	092LH8	092lh8 rhizobium m
15	564	26.5	098E16	098e16 rhizobium l
16	558	26.3	098X13	098x13 rickettsia
17	553	26.0	098X13	098x13 rickettsia
18	548	25.8	098X13	098x13 rickettsia
19	548	25.8	098X13	098x13 rickettsia
20	541.5	25.5	098X13	098x13 rickettsia
21	540.5	25.4	098X13	098x13 rickettsia
22	539.5	25.4	098X13	098x13 rickettsia
23	539.5	25.4	098X13	098x13 rickettsia
24	538.5	25.3	098X13	098x13 rickettsia
25	535.5	25.2	098X13	098x13 rickettsia
26	531.5	25.0	098X13	098x13 rickettsia
27	524	24.7	098X13	098x13 rickettsia
28	520	24.5	098X13	098x13 rickettsia
29	516.5	24.3	098X13	098x13 rickettsia
30	515.5	24.3	098X13	098x13 rickettsia
31	512	24.1	098X13	098x13 rickettsia

32	510.5	24.0	346	2	07M668	07m668 bordetella
33	510	24.0	339	2	06LV10	06lv10 photobacter
34	509.5	24.0	346	2	07W134	07w134 bordetella
35	507.5	23.9	357	2	089WV1	089wv1 bradyrhizob
36	503.5	23.7	341	2	07NUE3	07nue3 chromobacter
37	494.5	23.3	343	2	09CLJ1	09clj1 pasteurella
38	494	23.2	341	2	08ESI6	08esi6 oceanobacil
39	492.5	23.2	337	1	GCP_SALTY	P40731 salmonella
40	492.5	23.2	341	2	09ISV7	09isv7 pseudomonas
41	491	23.1	344	2	06FPY1	06fpy1 bartonella
42	490.5	23.1	337	2	08FD06	08fd06 escherichia
43	489	23.0	333	2	07VQ09	07vq09 candidatus
44	488.5	23.0	337	2	08XBR3	08xb3 escherichia
45	486.5	22.9	337	2	082XN2	082xn2 nitrosomona

ALIGNMENTS

RESULT 1
096EV9 PRELIMINARY; PRT; 414 AA.
AC 096EV9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE OSGBEP1 protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Stapleton M.J., Uedin T.B., Tothiyuki S., Carminci P., Prange C.,
RA Brownstein M.J., Uedin T.B., Tothiyuki S., Carminci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gutierrez P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyvinski M.I., Skalska U., Smalins D.B., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX Strausberg R.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: BC011904; AAH11904.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; P.O.-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; Zinc ion binding; IEA.
DR GO; GO:0006058; Proteolysis and peptidolysis; IEA.
DR InterPro; IPR000905; Peptidase_M22
DR InterPro; IPR009180; Pept_M22_Osai1gl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Csi1; 1.
DR PRINTS; PR00789; OSIALOPRASE-1.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRNAME; TIGR00329; gcp1_1.
DR SEQUENCE 414 AA; 45122 MW; A536B333F5C68BDD CRC64;

Query Match 100.0%; Score 2125; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.6e-155;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSKRYEFLRSFNFRPGTLFLHKIVLGIIETSCDDTAAAVVDETGVN 60
DB 1 MLILTKTAGVFPKSKRYEFLRSFNFRPGTLFLHKIVLGIIETSCDDTAAAVVDETGVN 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIRIVQEALSASGVSPSDLSAIAITTKPGL 120
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIRIVQEALSASGVSPSDLSAIAITTKPGL 120
QY 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTITLTNKEVPPVLIISGHCILAAV 180
DB 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTITLTNKEVPPVLIISGHCILAAV 180
QY 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLIKHPECSITSGKAIETHLAKQGNRFPHDIK 240
DB 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLIKHPECSITSGKAIETHLAKQGNRFPHDIK 240
QY 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVVR 300
DB 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVVR 300
QY 301 THRAILPCKORDLPONNAVIVASGVASNFYIRRALBITLTNAQCTLLCPPLCTDNG 360
DB 301 THRAILPCKORDLPONNAVIVASGVASNFYIRRALBITLTNAQCTLLCPPLCTDNG 360
QY 361 IMIANGIERLRAGIGLIHDIEGIRYEPKPLGVDISKEVGEASIKVPOLKMEI 414
DB 361 IMIANGIERLRAGIGLIHDIEGIRYEPKPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 2
Q9H4B0 PRELIMINARY; PRT; 439 AA.
AC Q9H4B0, -
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative sialoglycoprotease type 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Chen J.M., Fortunato M., Barrett A.J.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR MEROPS; M2.004; -
DR Genew; HGNC:23075; OSGEPL.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008223; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000905; Peptidase M2.
DR InterPro; IPR009180; Pept M22 Osi1gl.
DR Pfam; PF00814; Peptidase M22_1.
DR PIRSF; PIRSF004537; Osi1glc_ptide; 1.
DR PRINTS; PR00789; Osi1glc_ptide; 1.
DR PRODOM; PD002367; Peptidase M2.
DR TIGRPFAMS; TIGR00329; gcp; 1.
KM PROTEASE
SQ SEQUENCE 439 AA; 48040 MW; 44849372C784E41F CRC64;

Query Match 98.4%; Score 2090.5; DB 2; Length 439;
Best Local Similarity 93.8%; Pred. No. 1.3e-152;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLILTKTAGVFPKSKRYEFLRSFNFRPGTLFLHKIVLGIIETSCDDTAAAVVDETGVN 60

DB 1 MLILTKTAGVFPKSKRYEFLRSFNFRPGTLFLHKIVLGIIETSCDDTAAAVVDETGVN 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIRIVQEALSASGVSPSDLSAIAITTKPGL 120
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIRIVQEALSASGVSPSDLSAIAITTKPGL 120
QY 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTITLTNKEVPPVLIISGHCILAAV 180
DB 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTITLTNKEVPPVLIISGHCILAAV 180
QY 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLIKHPECSITSGKAIETHLAKQGNRFPHDIK 240
DB 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLIKHPECSITSGKAIETHLAKQGNRFPHDIK 240
QY 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVVR 300
DB 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVVR 300
QY 276 GQILSSAADIATVQHTMACHLVVRTHRALPCKORDLPONNAVIVASGVASNFYIR 335
DB 301 GQILSSAADIATVQHTMACHLVVRTHRALPCKORDLPONNAVIVASGVASNFYIR 360
QY 336 ALEILTNATQCTLLCPPLCTDNGIMANGIERLRAGIGLIHDIEGIRYEPKPLGVD 395
DB 361 ALEILTNATQCTLLCPPLCTDNGIMANGIERLRAGIGLIHDIEGIRYEPKPLGVD 420
QY 396 ISKEVGEASIKVPOLKMEI 414
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 3
Q96NH5 PRELIMINARY; PRT; 364 AA.
AC Q96NH5, -
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein FLJ30879.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimiya K., Ishihashi T., Yamashita T., Murakawa K., Fujimori K.,
RA Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Horuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara R., Tanase T., Nomura Y.,
RA Togiyasu S., Komai F., Hara R., Takeuchi K., Arita M., Iose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakabe H.,
RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitaira T.,
RA Ono T., Yamada K., Yamaki T., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki S., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuko Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Iosugi T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 cDNAs.";
 RL Nat. Genet. 36:40-45 (2004).
 DR EMBL; AK055441; BAB70923.1; -.
 DR MEROPS; M22.004; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009095; Peptidase M22.
 DR InterPro; IPR009180; Pept M22_Osialgl.
 DR Pfam; PF00814; Peptidase M22; 1.
 DR PIRSF; PIRSF04537; Osialglc_ptide; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KM Protease.
 SQ SEQUENCE 364 AA; 39528 MW; B0B605A07D0EC3D6 CRC64;
 Query Match 86.8%; Score 1845; DB 2; Length 364;
 Best Local Similarity 99.2%; Pred. No. 8,4e-134;
 Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MLITKTAGVFFPKSKRYVEFLSFNFHPTLFLHKIVLGITSCDDTAAAVVDEGNV 60
 DB 1 MLITKTAGVFFPKSKRYVEFLSFNFHPTLFLHKIVLGITSCDDTAAAVVDEGNV 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOQIHRENIQRIYQVQALSASGVSPSDLSAATTKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOQIHRENIQRIYQVQALSASGVSPSDLSAATTKPGL 120
 QY 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 DB 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 QY 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 DB 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 QY 181 QGVSDPFLLGSLDIAPGMDLVKVARSLTIHPECSTMSGKAIETHLAKQNRHFPDIK 240
 DB 181 QGVSDPFLLGSLDIAPGMDLVKVARSLTIHPECSTMSGKAIETHLAKQNRHFPDIK 240
 QY 241 PPLHAHAKNDSPFTGLQHTDKITMKKEKEGIEKQILSSAADIAATVQHTMACHLVR 300
 DB 241 PPLHAHAKNDSPFTGLQHTDKITMKKEKEGIEKQILSSAADIAATVQHTMACHLVR 300
 QY 301 THRALIFCKQKRDLPONNAVIVASGVASNFYIRRALBELITNAQCTLLCPPRLCTDNG 360
 DB 301 THRALIFCKQKRDLPONNAVIVASGVASNFYIRRALBELITNAQCTLLCPPRLCTDNG 360
 QY 361 IMIA 364
 DB 361 IMIA 364
 RESULT 4
 Q6PEB4 PRELIMINARY; PRT; 414 AA.
 AC Q6PEB4;
 DT 05-JUL-2004 (Tremblere, 27, Created)
 DT 05-JUL-2004 (Tremblere, 27, Last sequence update)
 DT 05-JUL-2004 (Tremblere, 27, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_TaxID=10090;
 RN
 RP SEQUENCE FROM N.A.
 RX STRAIN=CZECH II; TISSUE=Mammary tumor;
 RX MEDLINE=2238257; PubMed=1247932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkin R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Caesvant T.L., Schetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RA Strausberg R.;
 RT Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC058172; AAH58172.1; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009095; Peptidase M22.
 DR InterPro; IPR009180; Pept M22_Osialgl.
 DR Pfam; PF00814; Peptidase M22; 1.
 DR PIRSF; PIRSF04537; Osialglc_ptide; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBGAB CRC64;
 Query Match 86.4%; Score 1835; DB 2; Length 414;
 Best Local Similarity 85.0%; Pred. No. 5.8e-133;
 Matches 352; Conservative 25; Mismatches 137; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRYVEFLSFNFHPTLFLHKIVLGITSCDDTAAAVVDEGNV 60
 DB 1 MLITKTAGVFFPKSKRYVEFLSFNFHPTLFLHKIVLGITSCDDTAAAVVDEGNV 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOQIHRENIQRIYQVQALSASGVSPSDLSAATTKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOQIHRENIQRIYQVQALSASGVSPSDLSAATTKPGL 120
 QY 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 DB 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 QY 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 DB 121 ALSTGVLSFSLQVYQVKKEPIPIHMEAHALITRLTNKVEFPPLVLLISGCHLLAV 180
 QY 181 QGVSDPFLLGSLDIAPGMDLVKVARSLTIHPECSTMSGKAIETHLAKQNRHFPDIK 240
 DB 181 QGVSDPFLLGSLDIAPGMDLVKVARSLTIHPECSTMSGKAIETHLAKQNRHFPDIK 240
 QY 241 PPLHAHAKNDSPFTGLQHTDKITMKKEKEGIEKQILSSAADIAATVQHTMACHLVR 300
 DB 241 PPLHAHAKNDSPFTGLQHTDKITMKKEKEGIEKQILSSAADIAATVQHTMACHLVR 300
 QY 301 THRALIFCKQKRDLPONNAVIVASGVASNFYIRRALBELITNAQCTLLCPPRLCTDNG 360
 DB 301 THRALIFCKQKRDLPONNAVIVASGVASNFYIRRALBELITNAQCTLLCPPRLCTDNG 360
 QY 361 IMIANGIERLPAAGIILHDIGIYRPEKPCGVNISKVGVASIVPOLKKEI 414
 DB 361 IMIANGIERLPAAGIILHDIGIYRPEKPCGVNISKVGVASIVPOLKKEI 414
 RESULT 5
 Q8BLB6 PRELIMINARY; PRT; 414 AA.
 ID Q8BLB6

AC Q8BLB6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
 DE enriched library, clone:B230219017 product:similar to PUTATIVE
 DE SIALOGLYCOPROTEASE TYPE 2.
 GN Mus-Osegep11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RX STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RA The FANTOM Consortium;
 RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Kanno H., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Cornini F., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama T., Mishi K., Kitsuana T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hatada A.,
 RA Yamamoto K., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoaka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Ito M., Kagawa I., Kasukawa T.,
 RA Kurihara C., Kawai J., Kojima Y., Kondo S., Kono H., Kouda H., Koya S.,
 RA Kurihara C., Matuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito K., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AK045669; BAC32450.1; --

DR MEROPS; M22.004; --
 DR MGD; MG1.19335; Osegep11.
 DR GO; GO:0008450; P-O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009095; Peptidase M22.
 DR InterPro; IPR009180; Pept M22 Oe1a1gl.
 DR Pfam; PF00814; Peptidase_M22_1.
 DR PIRSF; PIRSF004537; Oe1a1glc-ptcds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;
 Query Match 86.0%; Score 1827; DB 2; Length 414;
 Best Local Similarity 85.0%; Pred. No. 2,4e-132;
 Matches 352; Conservative 23; Mismatches 39; Indels 0; Gaps 0;
 QY 1 MLILTKTAVFPFKPKSKRYEFLASFNHAPGTLFLKIVLGIETSCDDPTAAAVDETGVN 60
 DB 1 MLMLRTAGAIKPKPKSKYVGLRRSVHPRTLSCKLVGLIETSCDDTGAAVDETGVN 60
 QY 61 LGEALHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEALASAGVSPSDISAITIKPGL 120
 DB 61 LGEALHSQTEVHLKTGIVPPAAQOLHRENIQRIVEETISACRITPSDLSAITTIKPGI 120
 QY 121 ALSLGVSFSLSQVLQKKPFIPIHMEAAHLLTILNKVEPFLVLLISGCHLLALV 180
 DB 121 ALSLGVSFSLSQVLQKKPFIPIHMEAAHLLTILNKVEPFLVLLISGCHLLALV 180
 QY 181 QGVSDPLFLGKSLDIPAGMDKVARRLSLIKPECTSGGKAIHLAKQGRFHPDIK 240
 DB 181 QGVSDPLFLGKSLDIPAGMDKVARRLSLIKPECTSGGKAIHLAKQGRFHPDIK 240
 QY 241 PLHLAKNCDPSPFGVHTVDTKIIMKEKEEGIEKQIISADIAATVQHTVACHVVR 300
 DB 241 PPMQNNKNCDFSPFGVHTVDTKIITKEKEEGIEKQIISADIAATVQHTVACHVVR 300
 QY 301 TRAILFCQKQNLSPANNVAVVSGVANSVIRKALIEVANAQTCLCPRLCTDNG 360
 DB 301 TRAILFCQKQNLSPANNVAVVSGVANSVIRKALIEVANAQTCLCPRLCTDNG 360
 QY 361 IMIANNGIRLRAGIGIHDIGIRYEPKCPGVDSIKVGEASIVPOLKEI 414
 DB 361 IMIANNGIRLRAGIGIHDIGIRYEPKCPGVDSIKVGEASIVPOLKEI 414
 RESULT 6
 ID Q9DONO PRELIMINARY; PRT; 414 AA.
 AC Q9DONO;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 DE enriched library, clone:2610001M19 product:similar to PUTATIVE
 DE SIALOGLYCOPROTEASE TYPE 2.
 GN Name=Osegep11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RX STRAIN=C57BL/6J; TISSUE=Whole body;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.

1 MLTITKTAGVFFPKSRKRYEPLRSFNPHFGTLFLKIVIGIETSCDDTAAAVDBETGV 60

Db	1	MEMLRRTAGALPRPKSKVYGFRRRESVHRTTSSCHKLVGITSQDDTAAVVDETGV	6
Qy	61	LGSAHSQTEVHLKTGGIVPAAQOLRENIQRIQVEALSASGSPEDLSAATTKPGI	12
Db	61	LGSAHSQTEVHLKTGGIVPAAQOLRENIQRIQVEALSASGSPEDLSAATTKPGI	12
Qy	121	ALSLGVSLSLQVLQGLKKPFIPIHMEHAALTRITNKVEPFPVLVLSGGHCLALV	18
Db	121	ALSLGVSLSLQVLQGLKKPFIPIHMEHAALTRITNKVEPFPVLVLSGGHCLALV	18
Qy	181	QGVSDPFLILKSLDIAFGDMLDKVARRLSLIKHPEGSITWSGKAIEHLAKQNRFPFDIK	24
Db	181	QGVSDPFLILKSLDIAFGDMLDKVARRLSLIKHPEGSITWSGKAIEHLAKQNRFPFDIK	24
Qy	241	PELHAANCDPFSFTGLQHTVTDKTIIMKKKEKEGIEKGQILSSAADIAATVQHTACHLVKR	30
Db	241	PELHAANCDPFSFTGLQHTVTDKTIIMKKKEKEGIEKGQILSSAADIAATVQHTACHLVKR	30
Qy	301	THRAILFCRQDILLPQNNALVVASGVASFFYIRRALEITLTTAOTCTLLCPRLCTDNG	36
Db	301	THRAILFCRQDILLPQNNALVVASGVASFFYIRRALEITLTTAOTCTLLCPRLCTDNG	36
Qy	361	IMTAMNGIEELRAGLGIHTHIDIEGIRYEPKCPGLVDISKEVGEASIKVPOKHQEI	414
Db	361	IMTAMNGIEELRAGLGIHTHIDIEGIRYEPKCPGLVDISKEVGEASIKVPOKHQEI	414

RESULT 7

06AYN7 PRELIMINARY; PRT; 467 AA.

ID 06AYN7

AC 06AYN7

DT 25-OCT-2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Hypothetical protein (Fragment).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_Taxid=10116;

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Cassavani P.L., Scheetz T.E., Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C., Roha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J., Bosak S., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Harte S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs S.A., Fahey J., Helton E., Keteleman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzyzanski M.I., Shalaka U., Smolins D.B., Schmechel A., Schein J.B., Jones S.J., Maitra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX Director MGC Project;

RL Submitted (Aug-2004) to the EMBL/GenBank/DBJ databases.

EMBL: BC078974; AAT78974.1; -

GO: GO:0008450; P/O-staologycoprotein endopeptidase activity; IEA.

DR GO: GO:0008470; Zinc ion binding; IEA.

OR GO: GO:0006508; P/roteolysis and peptidolysis; IEA.

Query March		60.2%	Score 1279:	DB 2:	Length 404;
Best Local Similarity	64.1%;	Pred.	No. 3.9e-90;		
Matches 241;	Conservative 59;	Km.	Matches 76;	Indels 0;	Gaps 0
Dy	37 KIVIGIETSCDDTAAAVVDGTGNVGEAIIHSQTEVHLTKGGIVPPAAOQLHRENIQRITVO	96			
Dd	25 RLVIQIERISCDETGAADVDETRIRLIGESLSHQKHMTGCIIPLVAGRLHRENISRVOYQ	84			
Dy	97 EALSAGCVSPDDLAIATTIKPGLALSLGVGISFSLQVLVGKPKPIPHMEAHALTIR	156			
Dd	85 EALNRSAIEPSELTFVAATTVKRGIALSLGIGIDYSLKKEVRGHQKPFIPDHMEAHALTIR	144			
Dy	157 LTNKVEPEPLVLLISGCHCLATLVGVSDPFLILGSILDIAPEDMDIXVARPLSLIKHEEC	216			
Dd	145 MLHPIDEPFLVLVSGSHSLAALAGDIBFLLEQTLDABGDTLDTKIARLSTLNHHBEC	204			
Dy	217 STMSGGAKEIHIAKOGNRFHDIPKLPHARKCDFSFTGLGHVTDKIIMKKEKEGEBKG	276			
Dd	205 GTLSGGQAIERLAEGDGLAHFIISPQONVDNFSPAGLRQITGAILNKKEKEGEGVAG	264			
Dy	277 QILSSADIDIAATVOHTMACHLVKETHRAILFCRKORDLLPONNAVVLVASGVASNYIRRA	336			
Dd	265 QPLCSVKQIIAASQHTVASHAKRTHRAILFCCKSKGLLPENQPTLIVSGVASNEYIRQI	324			
Dy	337 LEILTMAQCULLCPRPRLCTDNNGMIMWNEIERBALGIHIDEGRIRYEPKCLGYDI	396			
Dd	325 LKITTDATGHLHCPSKFECTDNGVMIAWNGERLKQKGIILSYSEBVSEYBKAPLGADI	384			
Dy	397 SKEVGEASIKVPOLKM 412				
Dd	385 TSEVKEAIRKVPKCLK 400				
RESULT 10					
070918					
ID	PRELIMINARY;	PRT;	401 AA.		
AC	070918;				
DT	01-MAR-2004 (TREMBLrel. 26, Created)				
DT	01-MAR-2004 (TREMBLrel. 26, Last sequence update)				
DE	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
OS	AgCP14.990 (Fragment).				
OS	Name=agCG46164; ORFNames=EKSANGCG00000007922;				
OC	Anopheles gambiae str. PEST.				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pezomyia;				
OX	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.				
RN	NCBI_TaxID=180454;				
RP	[1]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=PEST;				
RU	Anopheles Genome Sequencing Consortium;				
CC	Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.				
CC	-!- CAUTION: The sequence shown here is derived from an				
CC	EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is				
CC	preliminary data.				
DR	EMBL; AAB01008900; EAA09387.1; -				
DR	GO; GO:00080450; F:O-sialoglycoprotein endopeptidase activity; IEA.				
DR	GO; GO:00080270; F:zinc ion binding; IEA.				
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.				
DR	InterPro; IPR009095; peptidase M22.				
DR	InterPro; IPR009180; pept_M22_Ostalg1.				
DR	Pfam; PF00814; Peptidase M22; 1.				
DR	PIRSF; PIRSFO04537; Ostalgic_ptds; 1.				
DR	PRINTS; PRO0789; OSIALOPTASB.				
DR	Prodom; PD002367; Peptidase M22; 1.				
DR	TIGRfams; TIGR00329; gcp; 1.				
FT	NON_TER				
SQ	SEQUENCE 401 AA; 43849 MW; CC9426723DJ3FD4P1 CRC64;				
Query March		26.4%	Score 772.5;	DB 2;	Length 401;
Best Local Similarity	42.7%;	Pred.	No. 4.2e-51;		
Matches 165;	Conservative 68;	Mismatches 142;	Indels 11;	Gaps 4;	

QY	38	YLVLGJETS	CDPDLAAAVNDETVGNVLGSAHISQTEVHLKTGQIVPPAAQOLHRENIQRI	VOE	97
Db	9	VVLGIETSCDDTLGALVLTGNGVTLGSEYHISQOSSHLPFGCIIIPPVADODIRANIESVQN			68
QY	98	ALSAAGVSPSDLSAATATTKIPGALSLVGVGLSFLQVLGQKKPFIPIHMEAHALTIRL			157
Db	69	APKLANMFPNDIDAVATNRPGPLPSLIVGMRYAKHIAHSYNNKPLPIHHMQHALLAMR			128
QY	158	TNKRVEPPLVLLISGGHCLLALVQGVSDLLAKSLDIAPGMDLVYARLSLKIHPBCS			217
Db	129	TSTIYPPFLCCLVVGSHSLVTFVESYARFRLIGETLDDAEGEALDXIKRLRNVAKYA			188
QY	218	TMSGKALIEHLAKQG----	NRPHPIKPLHAKKCDSPFGGLQVTDKIIIMKKKEBG		272
Db	189	QMSGGQALAEAAQQAQAGKDTSAIEPVL--	PLSKTRPCQSPALKNATATHLIERESTLH		246
QY	273	IEKGQILSSADIAATVQHTMACHLVXKTRTHRALPFCKORDLL--	PONNAVTVASGVASN		330
Db	247	LAPDALLDPYEAFCAFLKGVTRHMLHQRQRIEYERKRLPSDAPHSLVSGVACN			306
QY	331	FYIRPALLETINATQCTLLCPPLRLCTDNGIMIANNGIERLRA--	GLGILHDIETGRIYEP		388
Db	307	DVIFALSSMAAQFGSYTRPKKCLTDNGTMIANNGEKLAKOTAEMTTKYEQVDISG			366
QY	389	KCPGLGVDSKVEGSEASIKVPOUKMEI	414		
Db	367	KCPIGDSLIDVKEANIACKMAKAVDI	392		
RESULT 11					
Q9VWD6					
ID	Q9VWD6	PRELIMINARY;		PRT;	409 AA.
AC	Q9VWD6;				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)			
DE	CG14231-PA				
GN	ORENames=CG14231;				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;				
OC	Ephyroidae; Drosophilidae; Drosophila.				
OX	NCBI_Taxid=7227;				
EN	[1]				
EN	SEQUENCE FROM N.A.				
RP	MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;				
RA	Adams M.D., Gelinker S.B., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.H., Blasey R.G., Champe M., Pfeiffer B.D.,				
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,				
RA	Abriil J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,				
RA	Burtis K.C., Busam D.A., Butler C., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Foster C., Gabrielian A.B., Gang N.S., Gelbart W.M., Glasser K.,				
RA	Goddek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.H., Idegian C.,				
RA	Jatali M., Kalush F., Karpen G.H., Ke Z., Kenzison J.A., Ketchum K.A.,				
RA	Klimpel B.B., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai X.,				
RA	Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Lin X., Mattel P., McIntosh T.C., McLeod M.F., McPherson D.,				
RA	Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,				
RA	Mount S.M., Moy M., Murphy B., Murphy L.,				

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassenaar D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Ceiniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champagne M., Dugan S.P., Friese E., Hodgson A.,
 RA Paegel R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
 RA Paole J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Swirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
 RT "Finishing a whole-genome shotgun. Release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminler J.S., Bergman C.M., Krommiller B., Carlson J., Swirskas R.,
 RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Ceiniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomic perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Maira S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminler J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Ceiniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AEO03513; AAF49008.1; --
 DR Inact; O9VMD6; --
 DR FlyBase; FBgn0031060; CG14231.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PfamSF; PIRSF004537; Osialglc_Pptds; 1.
 DR PRINTS; PRK0789; OSIALOPTASE.
 DR PRODOM; PD002367; Peptidase_M22; 1.
 DR TRIPRAMS; TIGR00329; GCD; 1.
 DR SEQUENCE 409 AA; 45328 MW; 9797F66E7D155538 CRC64;
 QY Query Match 33.6%; Score 714.5; DB 2; Length 409;
 Best Local Similarity 42.0%; Pred. No. 1.3e-46;
 Matches 156; Conservative 60; Mismatches 140; Indels 15; Gaps 6;

DB 27 VGIETSCDDTGAIVDTGRTAVANLBSQGFHTYGGIIPPRADLRARIESAYQRC 86
 QY 99 LSASVSPEDLSAINTTKRGLASLGVLSLSLOVGOKKPFRTIHHMEANLRL 158
 DB 87 MSAALDKPQPLAIVATTPTRPGPLSLVGRPARHARLRQKPLVHHMEALDQARME 146
 QY 159 N-KYEPFLVLTIGSGHCLALVGVSPDFLLSGSLDAPGMDLVARRSLIKHPEC 216
 DB 147 HPEQIGPFLCLASGHCQLVANGRGRLTLTGOTLDAPRPAKIGRRRLRLHLPXY 206
 QY 217 STMSGKALIEHLAK-QGNRFHDIKPLHANNCFSTGLQHVYDKIIMKKEEGIEK 275
 DB 207 RLWNGRAIEHNAQLASDLAVEPFLPLAQQRNCNCSFAGIKNNSPRAIRAREARTPP 266
 QY 276 GQILSSADIATVQHTMACHVYKTRAILTC--KQRLPLEONNAVYVASGVASNPYI 333
 DB 267 DGVISNYGDFCAGLRSVSRHMRHRTORAIIEYCLPFRQLFGDPTPLVMSGVANNDAI 326
 QY 334 RRAELITVAQCCTLLCPPEPLCTDNGIMIANNGIERLRAGGILHDI-GIRYEPKCP 392
 DB 327 YANTHLLAAQYGRSFRPSKRYCSDNVMIAMHGVQOL-----LDKXASTRYID-- 377
 QY 393 GVDISKVGEA 403
 DB 378 SIDIGSAGFA 388
 QY 393 GVDISKVGEA 403
 DB 378 SIDIGSAGFA 388
 RESULT 12
 ID 022145 PRELIMINARY; PRT; 480 AA.
 AC 022145 OSBVL2;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoproteinase
 DE GCP1).
 GN Name:Ac2g45270; Synonyms=GCP1;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,
 RA Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hausenehl K., Adanaka I.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiyama A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Kimita M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,

RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narsaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Becker J.R.,
RA Theologis A.; to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002387; AAB82636.2; -
DR EMBL; AY024338; AAK00530.1; -
DR EMBL; AY063864; AAL36220.1; -
DR EMBL; AY117283; AAM51358.1; -
DR PIR; E84888; E84888.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptids; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRPFAMs; TIGR00329; gcp; 1.
KW Protease.
SQ SEQUENCE 480 AA; 5295 MW; 20DD6A86AC1FFAD CRC64;
Query Match 31.5%; Score 668.5; DB 2; Length 480;
Best Local Similarity 39.4%; Pred. No. 5.4e-43;
Matches 154; Conservative 60; Mismatches 130; Indels 47; Gaps 7;
QY 38 IVLGIEIETSCDDTAAYVDEGTNVLGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIYQ 97
DB 85 VVIGIEIETSCDDTAAYVVRNGEILISQVITSGQELLVQGVAPKQAEBAHSRVIDKVVQD 144
QY 98 ALSASGVSPDLSAATITIKGALSLGVGSLQVLGOLKKPPIPIHMEAAHLTR 157
DB 145 ALDKANITKEDLSAVAVYTIIGPGLSLCRVGVKARVAKRANSLPIVGHMEAAHLVRL 204
QY 158 T-NKVEPPLVLLISGGHCLTALVQVSDFLLGKSLDIAFGDMLDKVARRSLIKHPEC 216
DB 205 VEGELSPFPMALLISGGHCLTALVHLKLGQYQLGTVDDALGEAPDKAKMIGLGMH--- 261
QY 217 STMSGKAIEHLAKQGNRFHFDIKPPLHAKNCDSPFTGLQHTVDTKIMKEKEGIEKG 276
DB 262 --RSGGVAVEHLAEGDAKSKFVPMKHYKDCNPSYAGLKTQVRLALEANE-----IDAK 315
QY 277 QILSSA-----ADIAVQHTMACHLVKTRTHAILFCQKQDRLPQNNAVLVASGVA 328
DB 316 CPVSSATNEDRRNADIAASFQVAVLHLEKCEKRAIDMALE---LPSIKAMVIGGVA 372
QY 329 SNFYIRALBELTNAQCTLLCPPLRCTDNGIMIANNGIERLAPAGLILHDIIGIRYE- 387
DB 373 SNKYVRLNNIVENKQLKLVCPPLSCTDNGVAVVAMTGLHFRVG-----RYDP 422
QY 388 -----PKCPGLGVDSKEVGRA 403
DB 423 PPATPEPDVYVLDPRMPLEGEYAKGRSEA 453
RESULT 13
Q73H71 PRELIMINARY; PRT; 335 AA.
AC Q73H71;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, last annotation update)
DB Peptidase, M22 family protein.
GN OrderedLocuNames=WD0659;
OS Wolbachia pipiensis wmel;
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Wolbachieae; Wolbachia.
OX NCBI_TaxID=56077;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15024419;

RA Wu M., Sun L.V., Yamathayan J.J., Riegler M., Deboy R.T.,
RA Brownlie J.C., McGraw E.A., Martin W., Esser C., Ahmadshef N.,
RA Wiegand C., Madupu R., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,
RA Berry K.J., Young M.B., Utterback T.R., Weidman J.F., Nierman W.C.,
RA Paulsen I.T., Nelson K.E., Tettelin H., O'Neill S.L., Eisen J.A.;
RT "Phylogenomics of the reproductive parasite Wolbachia pipiensis wmel:
RT a streamlined genome overrun by mobile genetic elements.";
RL PLOS Biol. 2:327-341(2004).
DR EMBL; AB017258; AAS14395.1; -
DR TIGR; W00699; -
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptids; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRPFAMs; TIGR00329; gcp; 1.
KW Complete proteome.
SQ SEQUENCE 335 AA; 3634 MW; F4914CDA58BC9390 CRC64;
Query Match 27.2%; Score 577.5; DB 2; Length 335;
Best Local Similarity 36.3%; Pred. No. 3.6e-36;
Matches 128; Conservative 68; Mismatches 130; Indels 27; Gaps 6;
QY 37 KIVLGIEIETSCDDTAAYVDEGTNVLGEAHSQTEVHLKTGIVPPAAQQLHRENIQRIYQ 96
DB 2 KIVLAVETSCDEFAVAIVVSDQVLAHEILSQAE-HKGGGVLPBIAISRAHMEHLSGLIK 60
QY 97 EALSASGVSPDLSAATITIKGALSLGVGSLQVLGOLKKPPIPIHMEAAHLTR 156
DB 61 SAVERSNLNPCDMLNIAATSGPGLIGLIVGTMAVAIAHVAKPPIAVNHLTAHALVIR 120
QY 157 LNKVEPPLVLLISGGHCLTALVQVSDFLLGKSLDIAFGDMLDKVARRSLIKHPEC 216
DB 121 LKHEVPEPPLVLLISGGHCLTALVQVSDFLLGKSLDIAFGDMLDKVARRSLIKHPEC 174
QY 217 STMSGKAIEHLAKQGNRFHFDIKPPLHAKNCDSPFTGLQHTVDTKIMKEKEGIEKG 276
DB 175 -SYRGPPLIKELAKQGNRFRFLPRAMIRSGCNPSFSGIKTAVKVLVQELKRS----- 228
QY 277 QILSSAADIATVQHTMACHLVKTRTHAILFCQKQDRLPQNNAVLVASGVA SNFYIR 336
DB 229 ---QVVCVCA SFQECISDILDRVSNAILMAESLNKIND---FVITGVAAANNFLREK 282
QY 337 LEILTNAQCTLLCPPLRCTDNGIMIANNGIERLAPAGLILHDIIGIRYE 389
DB 283 LKQHN---LNFPPENDCTNAINVGMWTGIERLOKYN-----IDPLNFPAPR 327
RESULT 14
Q92LH8 PRELIMINARY; PRT; 360 AA.
AC Q92LH8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, last annotation update)
DB PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).
GN ORFNames=SWC03230;
OS Rickettsia meliocoli (Sinorhizobium meliocoli).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Battu J.,
RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kist B., Leilaure V., Masny D.,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 08:37:51 ; Search time 53.6902 Seconds
(without alignments)
1923.349 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALTIRLTKNKEPPLV.....DISKEVGASIKVQLKMEI 267

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	267	5 ABG96487	ABG96487 Novel hum
2	1385	100.0	414	5 ABG96478	ABG96478 Novel hum
3	1385	100.0	414	5 ABB05481	ABD05481 Human O-8
4	1385	100.0	414	6 ABJ26654	ABJ26654 Human pro
5	1348	97.3	414	5 AAE29234	AAE29234 Human gly
6	1348	97.3	414	6 ABG71161	ABG71161 Novel hum
7	1348	97.3	414	6 ABU09569	ABU09569 Human gly
8	1105	79.8	364	6 ADA54471	ADA54471 Human pro
9	415.5	30.0	409	4 ABB69133	ABB69133 Drosophill
10	409.5	29.6	463	3 AAY52216	AAY52216 Arabidops
11	398.5	28.8	443	3 AAG19287	AAG19287 Arabidops
12	398.5	28.8	439	3 AAG19286	AAG19286 Arabidops
13	398.5	28.8	444	3 AAE31054	AAE31054 Ehtlichia
14	342	24.7	251	5 AAE31054	AAE31054 Ehtlichia
15	321.5	23.2	312	6 ABU22934	ABU22934 Protein e
16	314.5	22.7	312	6 ABU16575	ABU16575 Protein e
17	308.5	22.3	350	6 ABUS5232	ABUS5232 Protein e
18	307	22.2	350	8 ADI05040	ADI05040 M. catarr
19	304.5	22.0	343	6 ABUS3303	ABUS3303 Protein e
20	291.5	21.0	341	4 AAU36205	AAU36205 Pseudomon
21	291.5	21.0	341	6 ABUS38276	ABUS38276 Pseudomon
22	291.5	21.0	241	7 AAG73342	AAG73342 P aerugin
23	291.5	21.0	401	7 ABO68626	ABO68626 Pseudomon
24	289.5	20.9	341	7 ADG73344	ADG73344 P aerugin
25	289.5	20.9	342	3 AAY52202	AAY52202 Haemophil

26	289.5	20.9	342	4 AAU35450	AAU35450 Haemophil
27	289.5	20.9	342	6 ABU30280	ABU30280 Protein e
28	289.5	20.3	340	6 ABU40514	ABU40514 Protein e
29	280.5	20.3	357	7 ADP06228	ADP06228 Bacterial
30	279.5	20.2	337	4 AAU38187	AAU38187 Salmonell
31	279.5	20.2	337	6 ABU47561	ABU47561 Protein e
32	279.5	20.2	337	6 ABUS0237	ABUS0237 Protein e
33	277.5	20.0	335	6 ABU27480	ABU27480 Protein e
34	277	20.0	309	6 ABU33614	ABU33614 Protein e
35	277	20.0	341	6 ABU40069	ABU40069 Protein e
36	276	19.9	421	5 ABG96491	ABG96491 Novel hum
37	275.5	19.9	337	4 AAU34711	AAU34711 E. coli c
38	275.5	19.9	337	6 ABU28771	ABU28771 Protein e
39	270.5	19.5	325	2 AAR26325	AAR26325 Glycerol
40	270.5	19.5	325	3 AAY52203	AAY52203 Pasteurel
41	270.5	19.5	337	3 AAY52204	AAY52204 Pasteurel
42	268.5	19.4	343	7 ABO62704	ABO62704 Klebsiell
43	266.5	19.2	338	6 ABM67812	ABM67812 Phototrab
44	263.5	19.0	354	6 ABU37844	ABU37844 Protein e
45	263	19.0	341	6 ABU41689	ABU41689 Protein e

ALIGNMENTS

RESULT 1
ID ABG96487 standard; protein, 267 AA.
AC ABG96487;
DT 11-DEC-2002 (first entry)
XX
DE Novel human metalloprotease MPI fragment #1.
XX
XX
KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder.
XX
OS Homo sapiens.
XX
PN WO200272751-A2.
XX
PD 19-SEP-2002.
XX
PF 05-FEB-2002; 2002WO-US003353.
PR 05-FEB-2001; 2001US-0266518P.
PR 10-APR-2001; 2001US-0282814P.
XX
PI (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Chen J, Feder J, Nelson TC, Duclos F, Kyrstek S,
XX WPI, 2002-723329/78.
XX N-PSDB; ABS76639.
XX
PT New isolated amino acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
XX neurological disorders.
XX
PS Claim 5; Page 29; 473pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or

CC ameliorating a medical condition, particularly an immune disorder, an
CC aberrant glutamate transport or motor neuron disorder, such as
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC condition. The compositions and methods are also useful for diagnosing,
CC prognosticating, treating, ameliorating and/or treating disorders
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC Alzheimer's disease or Parkinson's disease). This is the amino acid
CC sequence of a metalloprotease MPI protein
XX
SQ Sequence 267 AA;
Query Match 100.0%; Score 1385; DB 5; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.4e-147;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAHLTIRLTNRYEPFVLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLKVARR 60
DB 1 MEAHLTIRLTNRYEPFVLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLKVARR 60
QY 61 LSLIKHPECSWTSGKAIENHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTYDKIIMKK 120
DB 61 LSLIKHPECSWTSGKAIENHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTYDKIIMKK 120
QY 121 EKEEGIEKQILSSAADIAATVOHTMACHLVYRTRAILFCQKRDLLPNNNAVLAASGV 180
DB 121 EKEEGIEKQILSSAADIAATVOHTMACHLVYRTRAILFCQKRDLLPNNNAVLAASGV 180
QY 181 ASNFYIRRALIETLNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
DB 181 ASNFYIRRALIETLNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
QY 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267
DB 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267
DB 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267
RESULT 2
ABG96478 standard; protein; 414 AA.
ID ABG96478 standard; protein; 414 AA.
AC ABG96478;
XX
DT 11-DEC-2002 (first entry)
XX
DE Novel human metalloprotease MPI.
XX
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder.
XX
XX Homo sapiens.
XX OS
XX PN MO200272751-A2.
XX PD 19-SEP-2002.
XX PF 05-FEB-2002; 2002WO-US003353.
XX PR 05-FEB-2001; 2001US-0266518P.

PR 10-APR-2001; 2001US-0282814P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
PI MPI; 2002-723329/78.
DR N-PSDB; ABS76635.
XX
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
PR neurological disorders.
XX
XX Claim 5; Fig 1A-C; 47pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (i) encoding a
CC metalloprotease (MP-1). (i) is useful for preventing, treating, or
CC ameliorating a medical condition, particularly an immune disorder, an
CC aberrant glutamate transport or motor neuron disorder, such as
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC condition. The compositions and methods are also useful for diagnosing,
CC prognosticating, treating, ameliorating and/or treating disorders
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC Alzheimer's disease or Parkinson's disease). This is the amino acid
CC sequence of a metalloprotease MPI protein
XX
SQ Sequence 414 AA;
Query Match 100.0%; Score 1385; DB 5; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.8e-147;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAHLTIRLTNRYEPFVLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLKVARR 60
DB 1 MEAHLTIRLTNRYEPFVLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLKVARR 207
QY 61 LSLIKHPECSWTSGKAIENHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTYDKIIMKK 120
DB 208 LSLIKHPECSWTSGKAIENHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTYDKIIMKK 267
QY 121 EKEEGIEKQILSSAADIAATVOHTMACHLVYRTRAILFCQKRDLLPNNNAVLAASGV 180
DB 268 EKEEGIEKQILSSAADIAATVOHTMACHLVYRTRAILFCQKRDLLPNNNAVLAASGV 327
QY 181 ASNFYIRRALIETLNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
DB 328 ASNFYIRRALIETLNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 387
QY 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267
DB 388 PKCPLGVDSIKVGEASIKVPOLKMEI 414
RESULT 3
ABB05481 standard; protein; 414 AA.
ID ABB05481 standard; protein; 414 AA.
AC ABB05481;
XX
XX 19-APR-2002 (first entry)
XX
XX Human O-sialoglycoproteinase-like protein SEQ ID NO:2.
DE Human O-sialoglycoproteinase-like protein; OSGLP; enzyme.
XX
XX Human; O-sialoglycoproteinase-like protein; OSGLP; enzyme.
XX

OS Homo sapiens.
 XX CN1318550-A.
 XX
 PD 24-OCT-2001.
 XX
 PF 19-APR-2000; 2000CN-00106834.
 XX
 PR 19-APR-2000; 2000CN-00106834.
 XX
 PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
 XX
 PI Mao Y, Xie Y;
 XX
 WPI: 2002-115090/16.
 DR N-PSDB; ABA93268.
 XX
 PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
 XX for diagnosing, preventing and treating related diseases.
 PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.
 CC The present sequence represents human O-sialoglycoproteinase-like protein
 CC (OSGPLP). The present invention also describes: (1) the preparation of
 CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the
 CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP
 CC protein in screening its agonist, excitomotor and inhibitor and preparing
 CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP
 CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors
 CC and antibodies in treating diseases related to the abnormal OSGPLP gene
 CC and in preparing the medicine composite for the treatment
 XX
 SQ Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 5; Length 414;
 Best Local Similarity 100.0%; Pred. No. 2.8e-147;
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAALATRLTNKVEPPVLLISGHCILALVQGVSPFLIGKSLDIAFGMDLVARR 60
 DB 148 MEAALATRLTNKVEPPVLLISGHCILALVQGVSPFLIGKSLDIAFGMDLVARR 207
 QY 61 LSLIKHEPCSTMSGKATIEHLAKQGNRFHPIKPLHAKNDPSFTGQHYTDKIMKK 120
 DB 208 LSLIKHEPCSTMSGKATIEHLAKQGNRFHPIKPLHAKNDPSFTGQHYTDKIMKK 267
 QY 121 EKEGIEKQILSSAADIATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGV 180
 DB 268 EKEGIEKQILSSAADIATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGV 327
 QY 181 ASNFYIRALILTNATQCTLLCPPLCTDNGMIANGLIRLRLAGLIIHDIGIRYE 240
 DB 328 ASNFYIRALILTNATQCTLLCPPLCTDNGMIANGLIRLRLAGLIIHDIGIRYE 387
 QY 241 PKCPLGVDSKEVGASIKVPOLKMEI 267
 DB 388 PKCPLGVDSKEVGASIKVPOLKMEI 414

RESULT 4
 AB026654
 ID AB026654 standard; protein; 414 AA.
 XX
 AC AB026654;
 XX
 DT 01-MAY-2003 (first entry)
 DE Human protein modification + maintenance molecule protein SEQ ID No. 8.
 XX
 XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
 KM cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
 KM antitumor; hepatotropic; gynecological; antibacterial; virucide;
 KM protozoacide; antiparasitic; cell proliferative disease; PMOD;

KW protein modification and maintenance molecule; immunogenic fragment;
 KW cancer; autoimmune; inflammatory disease; neurological disorder;
 KW gastrointestinal; developmental; vesicle trafficking disorder; infection;
 KW protein-protein interaction; drug-target interaction;
 KW gene expression profile; human.

OS Homo sapiens.
 XX
 PN WO200300844-A2.
 XX
 PD 03-JAN-2003.
 XX
 PF 18-JUN-2002; 2002WO-US019360.
 XX
 PR 22-JUN-2001; 2001US-0300508P.
 XX
 PR 06-JUL-2001; 2001US-0303445P.
 PR 13-JUL-2001; 2001US-0303405P.
 PR 09-AUG-2001; 2001US-0311442P.
 PR 24-AUG-2001; 2001US-0314821P.
 PR 29-AUG-2001; 2001US-0315929P.
 PR 03-MAY-2002; 2002US-0378205P.

(INCY-) INCYTE GENOMICS INC.

Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM,
 Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee BA, Yue H,
 Foraythe LJ, Barroso I, Runkumar J, Griffin JA, Li JX, Yang J,
 Thangavelu K, Gietzen KJ, Ding L, Baugh MR, Borowsky ML, Yao MG,
 Walla NK, Maon PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK,
 Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebardjian Y,
 PI
 XX
 DR WPI: 2003-184039/18.
 DR N-PSDB; AET23207.

PT New isolated human PMOD polypeptide and polynucleotide, useful for
 PT diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
 PT infections.
 PS
 XX
 PS Claim 63; Page 182-183; 225pp; English.

CC The invention relates to an isolated polypeptide comprising: any of 28
 CC sequences of 48-1256 amino acids; a natural amino acid sequence at least
 CC 90% identical to the 28 amino acid sequences; 94% identical to a sequence
 CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
 CC acids, or 97% identical to a sequence of 242 amino acids, all given in
 CC the specification; or a biologically active or immunogenic fragment of
 CC the isolated polypeptide. The polypeptides and polynucleotides are useful
 CC in diagnosing, treating and preventing diseases or conditions associated
 CC with the decreased expression of protein modification and maintenance
 CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
 CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
 CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,
 CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
 CC endometriosis), developmental, vesicle trafficking disorders, and
 CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also
 CC useful in assessing the effects of exogenous compounds on the expression
 CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its
 CC fragments are useful in screening compounds for effectiveness as agonist
 CC or antagonist of the polypeptides, or in altering the expression of the
 CC target polynucleotide and compounds that specifically bind to or modulate
 CC the activity of the polypeptide. The microarray is useful in monitoring
 CC or measuring protein-protein interactions, drug-target interactions, and
 CC gene expression profiles. This sequence represents a human PMOD protein
 CC of the invention
 CC
 SQ Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 5; Length 414;
 Best Local Similarity 100.0%; Pred. No. 2.8e-147;
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAALATRLTNKVEPPVLLISGHCILALVQGVSPFLIGKSLDIAFGMDLVARR 60

```

Db      148 MEAHALTIRLTINKVFPPVLVLLISGGHCLALVQGVSDPLLKGSJDIAFGMDLKVARR 207
Qy      61 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNCDPSTGTGQHTYDKIIMK 120
Db      208 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNCDPSTGTGQHTYDKIIMK 267
Qy      121 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPQNNAVLVASGV 180
Db      268 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPQNNAVLVASGV 327
Qy      181 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 240
Db      328 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 387
Qy      241 PKCPILGVNISKVEGASIKVPOLKMEI 267
Db      388 PKCPILGVNISKVEGASIKVPOLKMEI 414

RESULT 5
AAE29234
ID      AAE29234 standard; protein; 414 AA.
XX
AC      AAE29234;
XX
DT      27-JAN-2003 (first entry)
XX
Db      Human glycoprotease 28472 protein.
XX
Kw      Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
Kw      rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;
Kw      hyperension; ischaemic heart disease; obesity; myocardial infarction;
Kw      endothelial cell disorder; Grave's disease; psoriasis; brain disorder;
Kw      Parkinson's disease; Alzheimer's disease; haematopoietic disorder;
Kw      cerebral oedema; metabolic disorder; liver disorder; platelet disorder;
Kw      chromosome mapping; tissue typing; gene therapy; neuroprotective;
Kw      cyrostatic; anorectic; cardiac; haemostatic.
XX
XX      Homo sapiens.
XX
FH      Key
FH      Location/Qualifiers
FT      1..108
FT      /note= "Non-transmembrane domain; N-terminal cytoplasmic
FT      domain"
FT      38..369
FT      /note= "Endopeptidase O-sialoglycoprotein hydrolase
FT      metalloprotease zinc glycoprotease sialoglycoprotease
FT      domain"
FT      109..132
FT      /note= "Transmembrane domain"
FT      133..164
FT      /note= "Non-transmembrane domain; non-cytoplasmic loop"
FT      138..152
FT      /note= "Glycoprotease domain"
FT      165..189
FT      /note= "Transmembrane domain"
FT      190..316
FT      /note= "Non-transmembrane domain; cytoplasmic domain"
FT      317..333
FT      /note= "Transmembrane domain"
FT      334..414
FT      /note= "Non-transmembrane domain"
FT      374..414
FT      /note= "Sialoglycoprotease type domain"
XX
PN      W0200274960-A2.
XX
PD      26-SEP-2002.
XX
PF      08-NOV-2001; 2001WO-US051427.
XX

```

```

PR      08-NOV-2000; 2000US-0246768P.
PR      08-NOV-2000; 2000US-0246772P.
PR      15-NOV-2000; 2000US-0249185P.
PA      (MILL-) MILLENNIUM PHARM INC.
PI      Leiby KR, Kapeller-Libermann R, Gluckmann M;
DR      MPI; 2002-759898/82.
DR      N-PSDB; AAD46856.
XX
XX      New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
XX      useful for diagnosing and treating cancer, immune, cardiovascular,
XX      hemotopoietic, brain, pain, metabolic, liver or platelet disorders, and
XX      in pharmacogenomics.
XX      Claim 1; Fig 8; 178pp; English.
XX
CC      The present invention relates to novel 38650, 28472, 5495, 65507, 81588
CC      or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
CC      protease or seven transmembrane domain (7TM) receptor family members.
CC      Sequences of the invention are useful in diagnosing and treating cancer
CC      or aberrant cellular proliferation and/or differentiation (e.g. colon or
CC      lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
CC      arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
CC      hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
CC      myocardial infarction, thrombus) including endothelial cell disorders,
CC      (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
CC      disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
CC      pain and metabolic disorders (e.g. obesity), liver disorders or platelet
CC      disorders. They are also useful in screening assays, predictive medicine
CC      (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
CC      and pharmacogenetics) and prophylactic and therapeutic methods. The
CC      nucleic acids may also be used in chromosome mapping, tissue typing and
CC      forensic biology and as surrogate markers. Sequences of the invention are
CC      also used in gene therapy. The present sequence is human glycoprotease
CC      28472 protein
XX
XX      Sequence 414 AA;
XX
Query Match      97.3%; Score 1348; DB 5; Length 414;
Best Local Similarity 97.4%; Pred. No. 4,3e-143;
Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Qy      1 MEAHALTIRLTINKVFPPVLVLLISGGHCLALVQGVSDPLLKGSJDIAFGMDLKVARR 60
Db      148 MEAHALTIRLTINKVFPPVLVLLISGGHCLALVQGVSDPLLKGSJDIAFGMDLKVARR 207
Qy      61 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNCDPSTGTGQHTYDKIIMK 120
Db      208 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNCDPSTGTGQHTYDKIIMK 267
Qy      121 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPQNNAVLVASGV 180
Db      268 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPQNNAVLVASGV 327
Qy      181 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 240
Db      328 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 387
Qy      241 PKCPILGVNISKVEGASIKVPOLKMEI 267
Db      388 PKCPILGVNISKVEGASIKVPOLKMEI 414

RESULT 6
ABG71161
ID      ABG71161 standard; protein; 414 AA.
AC      ABG71161;
XX
DT      30-JAN-2003 (first entry)
XX

```


DE Novel human glycoprotease 28472.
 XX
 KM Cancer; aberrant cell proliferation; aberrant cell differentiation;
 KM breast cancer; ovarian cancer; prostate cancer; colon cancer;
 KM lung cancer; immune disorder; heart disorder; cardiovascular disorder;
 KM endothelial disorder; hematopoietic disorder; blood vessel disorder;
 KM brain disorder; pain; metabolic disorder; liver disorder; diabetes;
 KM platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;
 KM autoimmune disorder; hypertension; atherosclerosis; heart failure;
 KM myocardial infarction; ischaemic heart disease; Crohn's disease;
 KM Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
 KM cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;
 KM Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT 138. .152
 FT Domain /label= Glycoprotease_domain
 XX
 PN W0200277233-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 08-NOV-2001; 2001WO-US046724.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Lei by KR, Kapeller-Libermann R, Glucksmann M;
 XX
 DR WPI; 2003-029938/02.
 DR N-PSDB; ABS57020.
 XX
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain
 PT nucleic acid and polypeptides, designated 38650, 28472, 5495, 65507,
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
 PT hypertension.
 XX
 PS Claim 4; Fig 8A-B; 178pp; English.
 XX
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
 CC sequences that encode a human seven transmembrane domain (7TM). The
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
 CC sequences are useful for diagnosing, preventing or treating a subject
 CC with or at risk of developing a disorder, e.g. cancer or aberrant
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
 CC prostate, colon or lung cancer), immune disorders, heart disorders,
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,
 CC liver disorders or platelet disorders. These disorders include carcinoma,
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
 CC cachexia or diabetes. This is the amino acid sequence of the novel human
 CC glycoprotease 28472
 XX
 XX Sequence 414 AA;
 Query Match 97.3%; Score 1348; DB 6; Length 414;
 Best Local Similarity 97.4%; Freq. No. 4.3e-143;
 Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

DB 148 MEAHALTITRLTNKVEPPVLVLLISGHCCLALVQGVSDFLLGKSLDIAPGMDLVKVAR 207
 QY 61 LSLIHPESTMSGGAIAIHLAKQGNRFHPDIKPRILAHAKNCFSTGLOHTWDKIMKK 120
 DB 208 LSLIKHPESTMSGGAIAIHLAKQGNRFHPDIKPRILAHAKNCFSTGLOHTWDKIMKK 267
 QY 121 EKEEGIEKQIISASADIATVQHTMACILVYKTRTHAIFCKORDLLPQNNAVLVASGGV 180
 DB 268 KQBEIEKQIISASADIATVQHTMACILVYKTRTHAIFCKORDLLPQNNAVLVASGGV 327
 QY 181 ASNFYIRRALBITLVNATQCTLLCPPERLCTDNGIMIMWNGIERLRAGLGIHDIEGIRYE 240
 DB 328 ASNFYIRRALBITLVNATQCTLLCPPERLCTDNGIMIMWNGIERLRAGLGIHDIEGIRYE 387
 QY 241 PKCPILGVDISKVEGASIVPQLKWEI 267
 DB 388 PKCPILGVDISKVEGASIVPQLKWEI 414
 RESULT 7
 AB009569
 ID AB009569 standard; protein; 414 AA.
 XX
 AC AB009569;
 XX
 DT 08-JUL-2003 (first entry)
 XX
 DE Human glycoprotease encoded by cDNA 28472.
 XX
 KW Human; enzyme; cancer; aberrant cellular proliferation; differentiation;
 KW immune disorders; heart disorder; brain disorder;
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;
 KW hematopoietic disorder; blood vessel disorder; metabolic disorder;
 KW liver disorder; platelet disorder; glycoprotease.
 XX
 OS Homo sapiens.
 XX
 PN US2003009017-A1.
 PN
 XX
 PD 09-JAN-2003.
 XX
 PF 08-NOV-2001; 2001US-00012140.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (LEIBY) LEIBY K R.
 PA (KAPL) KAPPELLER-LIBERMANN R.
 PA (GLUC/) GLUCKSMANN M A.
 XX
 PI Lei by KR, Kapeller-Libermann R, Glucksmann M;
 XX
 DR WPI; 2003-428888/40.
 DR N-PSDB; ACA60887.
 XX
 PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic
 PT and liver disorders.
 XX
 PS Claim 4; Fig 8; 90pp; English.
 XX
 CC The invention relates to an isolated 38650 (encoding adenosine
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
 CC a sequence which is at least 60% identical to the six nucleic acids or
 CC their open reading frames, fragments of at least 15 nucleotides,
 CC naturally occurring variants, or a DNA insert of the plasmid deposited
 CC with the American Type Culture Collection as Accession No. not defined in
 CC the specification, which encodes the amino acid sequence). Also included
 CC are a host cell containing the nucleic acids (used to produce the

CC proteins), the encoded proteins, an antibody that selectively binds to
CC the polypeptide, and identifying a compound that binds to/modulates the
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
CC methods are useful for diagnosing, treating cancer, aberrant cellular
CC proliferation and/or differentiation, immune disorders, heart disorders,
CC cardiovascular disorders including endothelial cell disorders,
CC hematopoietic disorders, blood vessel disorders, brain disorders, pain
CC and metabolic disorders, liver disorders and platelet disorders (many
CC examples of these disorders are given in the specification). The present
CC sequence is the Human glycoproteinase encoded by cDNA 28472

XX
SQ Sequence 414 AA;

Query Match 97.3%; Score 1348; DB 6; Length 414;

Best Local Similarity 97.4%; Pred. No. 4.3e-143; Indels 0; Gaps 0;

Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MEAAHATITRLTNKVEFPFLVLLISGHCILALVQGVSPFLIGKSLIDIPGMDLKVARR 60
DB 148 MEAAHATITRLTNKVEFPFLVLLISGHCILALVQGVSPFLIGKSLIDIPGMDLKVARR 207
QY 61 LSLIKHPECSTWSGKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKTIIMKK 120
DB 208 LSLIKHPECSTWSGKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKTIIMKK 267
QY 121 EKEBEGIEKQIILSSADIAATVQHTMACHLVYRTHRAILFCRKORDLLPNNNAVTVASGV 180
DB 268 EKEBEGIEKQIILSSADIAATVQHTMACHLVYRTHRAILFCRKORDLLPNNNAVTVASGV 327
QY 181 ASNFYIRALIEITLNATQCTLLCPPLCTDNGIMIANNGIERLPAIGLIDIGIRYE 240
DB 328 ASNFYIRALIEITLNATQCTLLCPPLCTDNGIMIANNGIERLPAIGLIDIGIRYE 387
QY 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
DB 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 8

ADA54471

ID ADA54471 standard; protein; 364 AA.

XX ADA54471;

XX 20-NOV-2003 (first entry)

XX Human protein, SEQ ID 2039.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

KW Gene Therapy; human; secretory protein; membrane proteins; cancer;

KW inflammatory disease; osteoporosis; neurological disease.

XX Homo sapiens.

XX EP1293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.

XX 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;

XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maehno Y;

XX WPI; 2003-395539/38.

XX N-PSDB; ADA52832.

PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.

XX Claim 14; SEQ ID NO 2039; 205pp; English.

PS The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.

XX
SQ Sequence 364 AA;

Query Match 79.8%; Score 1105; DB 6; Length 364;

Best Local Similarity 98.6%; Pred. No. 1.1e-115; Indels 0; Gaps 0;

Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEAAHATITRLTNKVEFPFLVLLISGHCILALVQGVSPFLIGKSLIDIPGMDLKVARR 60
DB 148 MEAAHATITRLTNKVEFPFLVLLISGHCILALVQGVSPFLIGKSLIDIPGMDLKVARR 207
QY 61 LSLIKHPECSTWSGKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKTIIMKK 120
DB 208 LSLIKHPECSTWSGKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKTIIMKK 267
QY 121 EKEBEGIEKQIILSSADIAATVQHTMACHLVYRTHRAILFCRKORDLLPNNNAVTVASGV 180
DB 268 EKEBEGIEKQIILSSADIAATVQHTMACHLVYRTHRAILFCRKORDLLPNNNAVTVASGV 327
QY 181 ASNFYIRALIEITLNATQCTLLCPPLCTDNGIMIA 217
DB 328 ASNFYIRALIEITLNATQCTLLCPPLCTDNGIMIA 364

RESULT 9

ABB69133

ID ABB69133 standard; protein; 409 AA.

XX ABB69133;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 34191.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL13236.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

XX Disclosure; SEQ ID NO 34191; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
 CC sequences (AB161840-AB16175) and the encoded proteins (AB57737-
 CC AB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 409 AA;

Query Match 30.0%; Score 415.5; DB 4; Length 409;
 Best Local Similarity 38.2%; Pred. No. 1.4e-37;

Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;

QY 1 MEAHALTRLTN--KVEFPVLVLLISGGHCLLALVQVSDFLLGKSLDIAPGMDLKVVA 58
 DB 136 MEAHALQARMEHPQIGYFPLCLLASGGHCOLVANNGRGLTLGQTLDDAPGEAFDKIG 195
 QY 59 RLSLIHPKPCSTMSGKAIIEHLAK-QGNRPHFDIKPRLHNAKCDSEFTGLQHTDKII 117
 DB 196 RLRLHLTPERLWNGRAIEHAAQLASDPLAYEPPLLAQQRNCFSPAGIKNNSPFAI 255
 QY 118 MKKEKEGIEKGQILSSADIAATVQHTMACHLVYKTRHATLFC--KORDLLPQNNAVIY 175
 DB 256 RAERBARTRPDGVISYNGDFCAGLSVSHLMKRTIRALBYCLLPHRQLPGDTPTPLV 315
 QY 176 ASGGVANSFYIRALEILTNAOTCTLLCPPRLCTDNGIMIANGIERTLAGLILHDIE 235
 DB 316 MSGGVANNDAIYANIEHLAAQYGRSPRSKRYCSDNVMIAMGVEGL-----LDQKE 369
 QY 236 -GIRYEPKPLGVDISKEVGEA 256
 DB 370 ASTRYDYD---SIDIQSAGFA 388

RESULT 10

AAV52216 standard; protein; 463 AA.

XX AAV52216; 2

XX 09-FEB-2000 (first entry)

XX Arabidopsis thaliana yjld protein homologue.

XX yjld protein; essential; Gram positive; Gram negative; conserved; motif;
 KW identification; antagonist; antibacterial; antibiotic; broad spectrum;
 KM treatment; infection; resistance; drug target.

XX Arabidopsis thaliana.

XX Location/Qualifiers

FT Region 86..96 /note="yjld conserved motif 3"

FT Region 111..131 /note="yjld conserved motif 4"

FT Region 152..198 /note="yjld conserved motif 2"

FT Region 208..259 /note="yjld conserved motif 1"

XX MO9954470-A2.

XX 28-OCT-1999.

XX 20-APR-1999; 99WO-BP002635.

XX 22-APR-1999; 99GB-00009423.

XX (GLAXO) GLAXO GROUP LTD.

XX

PI Arizoni F, Edgerton MD, Loferer H, Peitsch MC;
 XX WPI, 2000-013253/01.

XX Novel bacterial polypeptides used to identify broad spectrum antibiotics.

XX Claim 1; Fig 1; 55pp; English.

CC Proteins AAV52202-Y52217 encompass a novel family of proteins designated
 CC the yjld family, after the name given to the *Escherichia coli* family
 CC member. These proteins are essential for the survival of both Gram
 CC negative and Gram positive bacteria, although no function has as yet been
 CC ascribed to these proteins. The yjld proteins, fragments of yjld proteins
 CC (for example, fragments encompassing one or more conserved yjld motifs
 CC such as AAV52218-Y52284) and nucleotides encoding them can be used to
 CC identify antagonists and broad spectrum antibacterial compounds. These
 CC antagonists and compounds can be used to treat a wide range of bacterial
 CC infections. New antibiotics are urgently needed, as serious bacterial
 CC infections and antibiotic resistant strains are becoming increasingly
 CC prevalent. The proteins of the invention are essential proteins for
 CC bacterial viability, and represent new targets for antibiotics

XX Sequence 463 AA;

Query Match 29.6%; Score 409.5; DB 3; Length 463;
 Best Local Similarity 36.6%; Pred. No. 8.2e-37;

Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;

QY 1 MEAHALTRLTN--KVEFPVLVLLISGGHCLLALVQVSDFLLGKSLDIAPGMDLKVVA 59
 DB 192 MEAHALVRLVQBELSPFMALLISGHNLVLAKHGLGYTLGTLVDALIEAPDKTK 251
 QY 60 RLSLIHPKPCSTMSGKAIIEHLAKQGNRPHFDIKPRLHNAKCDSEFTGLQHTDKII 119
 DB 252 WLGLDMH-----RSGGPAVEBELALEGDAKSVFNPMPMKHKNCFNSYAGLKTQVLAIRA 306
 QY 120 KEKEGIEKGQILSSADIAATVQHTMACHLVYKTRHATLFC--KORDLLPQNNAVIY 179
 DB 307 KE-----IRRADIASFORVAVILHEKGERALDMLR--LEPSIKHMYISGG 353
 QY 180 VASNFYIRALEILTNAOTCTLLCPPRLCTDNGIMIANGIERTLAGLILHDIEGIRY 239
 DB 354 VASNRYVRLNNTVENKRLKLVCPPLSLCTDNGVAVMTGLEHRRVGS-----RY 403
 QY 240 E-----KCPPLGVDISKEVGEA 256
 DB 404 DPPPATPEPDVYDLRPRMPLGSEYAKGRSBA 436

RESULT 11

AAG19287 standard; protein; 245 AA.

XX AAG19287;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 21031.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000BP-00301439.

XX 25-FEB-1999; 99US-0121825P.
 XX 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130448P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137724P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149922P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.

PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 245;
Best Local Similarity 38.7%; Pred. No. 5.6e-36;
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAAATLIRLT-ANVEPPFLVLLISGGHCLALVGVSDPFLIGSLDIARQDMIDKAR 59
DB 1 MEAAATLIRLT-ANVEPPFLVLLISGGHCLALVGVSDPFLIGSLDIARQDMIDKAR 60
QY 60 RLSTIKPECSITMSGKAIIEHLAKQNRFFHDIKPILHAKNCDPSFPGLOHVTDKIMK 119
DB 61 WLGLDMH-----RSGGPAVEELALEGDAKSVKFNPMKYNKHCNFSYGLKQVRLAIEA 115
QY 120 KEKEEGIEKQQLISA-----ADIAATVOHTWACHLVKTRHAILFCQQRDLLPQNN 171
DB 116 KE-----IDAKCPVSSATNEDERNRADIASFQVAVLHBEKCEAIDMALE---LBPSI 168
QY 172 AVLVASGVASNFYIRALLETILTAOTCTLLCPPRILCTDNGIMTANNGIERLAGLIL 231
DB 169 KHWISGVASNKKYRRLNNIVENKNILKVCPPSLCTDNGVAVMGLBHFHVG----- 224
QY 232 HDIBGIRYEPKCP 244
DB 225 -----RYDEPPP 231

RESULT 12
ID AAG19286 standard; protein: 439 AA.

XX AAG19286;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21030.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 06-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138647P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.
XX 18-JUN-1999; 99US-0139461P.
XX 18-JUN-1999; 99US-0139462P.
XX 18-JUN-1999; 99US-0139463P.
XX 18-JUN-1999; 99US-0139750P.
XX 18-JUN-1999; 99US-0139763P.
XX 21-JUN-1999; 99US-0139817P.
XX 22-JUN-1999; 99US-0139899P.
XX 23-JUN-1999; 99US-0140353P.
XX 23-JUN-1999; 99US-0140354P.
XX 24-JUN-1999; 99US-0140695P.
XX 28-JUN-1999; 99US-0140823P.
XX 29-JUN-1999; 99US-0140991P.
XX 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 25-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.

PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161932P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 439;
Best Local Similarity 38.7%; Pred. No. 1,3e-35;
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAAALTRLT-NKVEPPPLVLLISGGHCLLALVQGVSDPFLLLGSLDIAPGDMLDKVAR 59
DB 195 MEAHTLVARLVQGEISFPFMALLISGHNVLVIAHKLGQYTGTVDDAIGSAFDTAK 254
QY 60 RLSLTKHPECSTMSGKAIIEHLAKQGNRHFPIKPPHLHAKNCDSFTGLQHYTDKTIWK 119
DB 255 WLGLDMH-----RSGGPAVEELALBEGDAKSVENPMKXKCNFVAGLKTQVRLAIFA 309
QY 120 KEKEGIEKGLISGA-----ADIAATVOHTMACHLVKTRHRIILFCXKORDLLPNN 171
DB 310 KE-----IDAKCPVSSATNEDRRNRADIASFORAVAHLEEKCEKRIADALE---LEPSI 362
QY 172 AVLVASGVASNFIYRRALEILITNATOCTLLCPPRPLCTDNGIMIAMNGIEIRLRAIGIL 231
DB 363 KHMVISGGVASNKYRRLRNINVENKNLKLVCPPSLCTDNGVMVMTGLBHRVG----- 418
QY 232 HDIEGIRYBPCKP 244
DB 419 -----RYDPPPP 425

RESULT 13
AAG19285
ID AAG19285 standard; protein; 444 AA.
XX
AC AAG19285;
XX

DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21029.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147415P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.


```
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156566P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 444;
Best Local Similarity 38.7%; Pred. No. 1.4e-35;
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAAHTIRLTKNKEPPFLVLIISGGHCLATVGVSDPDLIGKSIDIAPGMDLKYAR 59
DB 200 MEAHTIRLTKNKEPPFLVLIISGGHCLATVGVSDPDLIGKSIDIAPGMDLKYAR 259
QY 60 RLSLIKRPECSTWGGKAIIEHLAKOGNRFHFDIKPPLHAKNCDPSTGLQHYVTDKIMK 119
DB 260 WLGIDMH-----RSGGPAVEELALEGDAKSVKFNPMKYHDCNFSYAGLKTQVLAIEA 314
QY 120 KEKEBGEIKQIILSSA-----ADIAATVOHTMACHLVKRTTRAILFCQKRDLLPQNN 171
DB 315 KE-----IDAKCPVSAETMEDRRNRADIAASFQRAVAVLHLEEKCEKRAIDWALB---LEPSI 367
QY 172 AVIVASGVASNFYIRALBILTNATQCTLLCPPLCTDNGIMIAWNGIERLRAGLGL 231
DB 368 KHWYISGVASNFYIRALBILTNATQCTLLCPPLCTDNGIMIAWNGIERLRAGLGL 423
QY 232 HDIGIRYEPKCP 244
```

```
DB 424 -----RYDPPP 430

RESULT 14
AAE31054
ID AAE31054 standard; protein; 251 AA.
AC AAE31054;
XX
XX
DT 24-FEB-2003 (first entry)
XX
XX Ehrlichia ruminantium hypothetical o-sialoglycoprotein endopeptidase.
XX
XX Vaccine; immunity; rickettsial infection; spotted fever; heart water;
XX tynus; pathogen; immunostimulant; antibacterial; lhwoif4; enzyme;
XX o-sialoglycoprotein endopeptidase.
XX Ehrlichia ruminantium.
XX
XX MO20026652-A2.
XX
XX 29-AUG-2002.
XX
XX 20-FEB-2002; 2002MO-US005772.
XX
XX 20-FEB-2001; 2001US-0269944P.
XX
XX (UYFL ) UNITV FLORIDA.
XX
XX Barbet AF, Whitmire WW, Kamper SM, Simbi BH, Ganta RR;
XX Moreland AL, Mwangi DM, McGuire TC, Mahan SM;
XX MPI: 2002-723186/78.
XX N-PSDB; AAD48239.
XX
XX New Ehrlichia ruminantium polynucleotides, useful as vaccines for
XX inducing protective immunity, and protecting animals or humans against
XX rickettsial diseases, e.g. tynus, spotted fever or heart water.
XX
XX Claim 8; Page 67-68; 206pp; English.
XX
XX The present invention relates to nucleic acid vaccines for conferring
XX immunity to rickettsial infection, including Ehrlichia ruminantium
XX (formerly Cowdria ruminantium). The invention also relates to novel B.
XX ruminantium polynucleotides and their corresponding proteins. Sequences
XX of the invention are useful for inducing immunity, particularly
XX protective immunity. They are also useful for detecting the presence of
XX B. ruminantium in a biological sample. They are useful in vaccines for
XX protecting animals or humans against rickettsial diseases, e.g. tynus,
XX spotted fever or heart water. Sequences of the invention are useful for
XX detecting antibodies to pathogens. The present sequence is B. ruminantium
XX hypothetical o-sialoglycoprotein endopeptidase (lhwoif4) protein
XX
XX Sequence 251 AA;

Query Match 24.7%; Score 342; DB 5; Length 251;
Best Local Similarity 34.8%; Pred. No. 1.4e-29;
Matches 86; Conservative 39; Mismatches 90; Indels 32; Gaps 6;

QY 1 MEAAHTIRLTKNKEPPFLVLIISGGHCLATVGVSDPDLIGKSIDIAPGMDLKYAR 60
DB 20 LEAHTIRLTKNKEPPFLVLIISGGHCLATVGVSDPDLIGKSIDIAPGMDLKYAR 79
QY 61 ISLIRKPECSTWGGKAIIEHLAKOGNRFHFDIKPPLHAKNCDPSTGLQHYVTDKIMK 120
DB 80 ISLIRKPECSTWGGKAIIEHLAKOGNRFHFDIKPPLHAKNCDPSTGLQHYVTDKIMK 132
QY 121 KEKEBGEIKQIILSSAADIAATVOHTMACHLVKRTTRAILFCQKRDLLPQNNAVY 175
DB 133 KYIDN-----DFICNISASPODCIGDILVIRITNAIMSAINKINK-----LV 177
QY 176 ASGVASNFYIRALBILTNATQCTLLCPPLCTDNGIMIAWNGIERLRAGLGLHDE 235
```

Db 178 VTGVAHAHLRNRSICKONNFEVLYPETELCTDNGIMGWAGIENTSKGY-----VS 232

QY 236 GIRYEPK 242
 : : ||

Db 233 NLDFVPK 239

RESULT 15

ID	ABU22934	standard; protein; 382 AA.
xx		

AC ABU22934;

DT 19-JUN-2003 (first entry)
 YY

DE Protein encoded by Prokaryotic essential gene #8461..

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Bordetella pertussis.

PN WO200277183-A2.

PD 03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.
05-SEP-2001; 2001US-00848993
PB

PR 25-OCT-2001; 2001US-0342923P.
09 FEB 2002 0000TTC 00073957

PR 06-MAR-2002; 2002US-0362699P.
XX

PA (ELIT-) ELITRA PHARM INC.
XX

PI wang L,	Zamudio C,	Malone C
PI wang D,	Trawick JD	Carr GT

XX
XX
WPT: 2003-029926/02

DK N-PSDB; ALCA26804.
XX

PI new antisense nucle
PT for homologous nucle

XX F1. isolate candidate inc

XX
CS
CJALM 23; 56Q 1D NO

the 6213 antisense

CC (1) a vector comprising

CC nucleic acid; (2) a

antisense nucleic ac

CC proliferation; or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-regulated gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 382 AA;

Query Match	23.2%	Score 321.5	DB 6	Length 382
Best Local Similarity	37.4%	Pred. No. 5.6e-27		
Matches 88	Conservative 38	Mismatches 86	Indels 21	Gaps 8

[illegible]

Search completed: February 16, 2005, 13:05:56
Job time : 55.6902 secs

GenCore version 5.1.6 -
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:43:41 ; Search time 15.962 Seconds
(without alignments)
1248.675 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALTRLTNKKVPEFLV.....DISKEVGASIKVPOLKMEI 267

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	267	US-10-067-443-22	Sequence 22, Appl
2	1385	100.0	414	US-10-067-443-2	Sequence 2, Appl
3	1358.5	98.1	439	US-10-067-443-19	Sequence 19, Appl
4	409.5	29.6	463	US-10-067-443-3	Sequence 3, Appl
5	307	22.2	350	US-09-540-236-2726	Sequence 2726, Ap
6	291.5	21.0	401	US-09-252-991A-17372	Sequence 17372, A
7	280.5	20.3	357	US-09-543-681A-6513	Sequence 6513, Ap
8	279	20.1	342	US-08-087-797-3	Sequence 3, Appl
9	276	19.9	421	US-10-067-443-4	Sequence 4, Appl
10	276	19.9	421	US-10-067-443-28	Sequence 28, Appl
11	270.5	19.5	325	US-08-087-797-2	Sequence 2, Appl
12	268.5	19.4	343	US-09-489-039A-9221	Sequence 9221, Ap
13	261	18.8	363	US-09-107-532A-6609	Sequence 6609, Ap
14	254	18.3	336	US-08-987-121A-4	Sequence 4, Appl
15	250	18.1	335	US-08-961-083-52	Sequence 52, Appl
16	250	18.1	335	US-09-536-784-52	Sequence 52, Appl
17	248	17.9	336	US-09-066-512-2	Sequence 52, Appl
18	246	17.8	336	US-09-583-110-4857	Sequence 4857, Ap
19	243	17.5	327	US-10-067-443-5	Sequence 5, Appl
20	241	17.4	336	US-09-107-433-4221	Sequence 4221, Ap
21	236	17.0	272	US-09-710-279-728	Sequence 728, App
22	236	17.0	366	US-09-134-000C-4956	Sequence 4956, Ap
23	233	16.8	368	US-09-134-001C-3909	Sequence 3909, Ap
24	220	15.9	341	US-09-145-624-2	Sequence 2, Appl
25	218	15.7	344	US-09-198-452A-213	Sequence 213, Appl
26	218	15.7	360	US-09-438-185A-196	Sequence 196, App
27	211.5	15.3	344	US-09-602-777A-148	Sequence 148, App

28	194	14.0	340	US-10-067-443-6	Sequence 6, Appl
29	181.5	13.1	143	US-09-328-352-4387	Sequence 4387, Ap
30	169	12.2	292	US-09-724-623-81	Sequence 81, Appl
31	157	11.3	214	US-09-328-352-4609	Sequence 4609, Ap
32	94.5	6.8	480	US-09-583-110-5050	Sequence 5050, Ap
33	94.5	6.8	481	US-09-107-433-3197	Sequence 3197, Ap
34	83	6.0	42	US-09-902-540-13841	Sequence 13841, A
35	81.5	5.9	1087	US-08-264-002-5	Sequence 5, Appl
36	81	5.8	1072	US-09-902-540-15572	Sequence 15572, A
37	79	5.7	328	US-09-710-279-3008	Sequence 3008, Ap
38	79	5.7	336	US-09-134-001C-5549	Sequence 5549, Ap
39	79	5.7	445	US-08-083-945C-2	Sequence 2, Appl
40	79	5.7	445	US-08-083-945C-7	Sequence 7, Appl
41	79	5.7	578	US-09-252-991A-31318	Sequence 31318, A
42	79	5.7	1137	US-09-538-092-968	Sequence 968, App
43	78.5	5.7	503	US-09-252-991A-22790	Sequence 22790, A
44	78.5	5.7	1658	US-08-609-049A-13	Sequence 13, Appl
45	78.5	5.7	1658	US-09-170-996-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1					
US-10-067-443-22					
Sequence 22, Application US/10067443					
Patent No. 6642041					
GENERAL INFORMATION:					
APPLICANT: Bristol-Myers Squibb Company					
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN					
FILE OF INVENTION: SPINL CORD, ME-1					
FILE REFERENCE: D0073 NP					
CURRENT APPLICATION NUMBER: US/10/067,443					
CURRENT FILING DATE: 2002-02-05					
PRIOR APPLICATION NUMBER: US 60/266,518					
PRIOR FILING DATE: 2001-02-05					
PRIOR APPLICATION NUMBER: US 60/282,814					
PRIOR FILING DATE: 2001-04-10					
NUMBER OF SEQ ID NOS: 71					
SOFTWARE: PatentIn version 3.0					
SEQ ID NO 22					
LENGTH: 267					
TYPE: PRT					
ORGANISM: homo sapiens					
US-10-067-443-22					
Query Match					
Best Local Similarity 100.0%; Pred. No. 3.6e-160;					
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MEAHALTRLTNKKVPEFLVLLISGGHCLALVGVSPFLIGKSLDAPGMDLKVARR	60		
DB	1	MEAHALTRLTNKKVPEFLVLLISGGHCLALVGVSPFLIGKSLDAPGMDLKVARR	60		
QY	61	LSLTHGPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGQHTYTDKIMKK	120		
DB	61	LSLTHGPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGQHTYTDKIMKK	120		
QY	121	EKEEGIEKGQILSSADIAATVOHTMAGLVKTRTALIFCKQRODLLPONNAVLAASGV	180		
DB	121	EKEEGIEKGQILSSADIAATVOHTMAGLVKTRTALIFCKQRODLLPONNAVLAASGV	180		
QY	181	ASNFTYRALBILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGLIDIGIRYE	240		
DB	181	ASNFTYRALBILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGLIDIGIRYE	240		
QY	241	PKCPGVDSIKVGEASIKVPOLKMEI	267		
DB	241	PKCPGVDSIKVGEASIKVPOLKMEI	267		
RESULT 2					
US-10-067-443-2					

Sequence 2, Application US/10067443
Patent No. 6642041
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: Patentin version 3.0
SEQ ID NO 2
LENGTH: 414
TYPE: PRT
ORGANISM: Homo sapiens
US-10-067-443-2

Query Match 100.0%; Score 1385; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 7.3e-160;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MEAHLATRLTNKVEFPPLVLLISGHCILALVQGVSPFLIGKSLDIAPGMDLKVARR 60
148 MEAHLATRLTNKVEFPPLVLLISGHCILALVQGVSPFLIGKSLDIAPGMDLKVARR 207
61 LSLIKHEPCSTWSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKIIMKK 120
208 LSLIKHEPCSTWSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKIIMKK 267
121 EKEEGIKGQILSSAADIATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGV 180
268 EKEEGIKGQILSSAADIATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGV 327
181 ASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGGILHDIGIRY 240
328 ASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGGILHDIGIRY 387
241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
388 PKCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 3
US-10-067-443-19
Sequence 19, Application US/10067443
Patent No. 6642041
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: Patentin version 3.0
SEQ ID NO 19
LENGTH: 439
TYPE: PRT
ORGANISM: Homo sapiens
US-10-067-443-19

Query Match 98.1%; Score 1358.5; DB 4; Length 439;
Best Local Similarity 91.1%; Pred. No. 1.4e-156;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 MEAHLATRLTNKVEFPPLVLLISGHCILALVQGVSPFLIGKSLDIAPGMDLKVARR 60
DB 148 MEAHLATRLTNKVEFPPLVLLISGHCILALVQGVSPFLIGKSLDIAPGMDLKVARR 207
QY 61 LSLIKHEPCSTWSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKIIMKK 120
DB 208 LSLIKHEPCSTWSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKIIMKK 267
QY 121 EKEEGIKGQILSSAADIATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGV 180
DB 268 EKEEGIKGQILSSAADIATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGV 327
QY 181 ASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGGILHDIGIRY 240
DB 328 ASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGGILHDIGIRY 387
QY 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
DB 388 PKCPLGVDISKEVGEASIKVPOLKMEI 439

RESULT 4
US-10-067-443-3
Sequence 3, Application US/10067443
Patent No. 6642041
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: Patentin version 3.0
SEQ ID NO 3
LENGTH: 463
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-067-443-3

Query Match 29.6%; Score 409.5; DB 4; Length 463;
Best Local Similarity 36.6%; Pred. No. 8.2e-41;
Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;

QY 1 MEAHLATRLTNKVEFPPLVLLISGHCILALVQGVSPFLIGKSLDIAPGMDLKVARR 59
DB 192 MEAHLATRLTNKVEFPPLVLLISGHCILALVQGVSPFLIGKSLDIAPGMDLKVARR 251
QY 60 LSLIKHEPCSTWSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKIIMKK 119
DB 252 WLGUDMH-----RSSGPAVEBELALEGDAKSVFENVPMKHKDCNSYAGLKTQYVALIA 306
QY 120 EKEEGIKGQILSSAADIATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGV 179
DB 307 KE-----IRNRADIAASFQVAVVHLERKCEKRAIDMALE---LEPSIKHWISGG 353
QY 180 ASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGGILHDIGIRY 239
DB 354 VASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGGILHDIGIRY 403
QY 240 E-----PKCPLGVDISKEVGEA 256
DB 404 DEPPRATRPEDVYDLRPRWPLGEEVAKGRSRA 436

RESULT 5
US-09-540-236-2726
Sequence 2726, Application US/09540236
Patent No. 6673910

```

; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2726
; LENGTH: 350
; TYPE: PRT
; ORGANISM: M. catarrhalis
; US-09-540-236-2726

```

Query Match 22.2%; Score 307; DB 4; Length 350;

Best Local Similarity 33.3%; Pred. No. 1.7e-28;

Matches 84; Conservative 37; Mismatches 91; Indels 40; Gaps 7;

```

QY 1 MEAAATLIRLT--NKVEPFLVLLISGGHCLALVQVSDPFLLGKSLDIAFGDMLDKVA 58
DB 119 MEGHLLAPLADDPSPFPVCLVSGGHTLVADGCVYQILGESIDDAVGECPDKTA 178
QY 59 RLSLIKHPECSTMSGKAIHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTVDKIM 118
DB 179 KMLGL-DYP-----GGPNIEKLAKNGNPHAYELPRPMQH-KGLDFSPGKMTALHNLIX 230
QY 119 KEKEBGEIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKQDLPONNAVVASG 178
DB 231 DTPNAQSDP-----ATRADIASFEYAVVDLVKKCTALQMTGIRQ-----LVVAG 277
QY 179 GVASNFYIRRALIEILTNATQCTLLCPPLCTDNGIMIANNGIBRLRAG----- 227
DB 278 GVSANQIRLRILTTLAQIDASVYAPTELCTDNGAMIAVGFCHLSKGSDDLAVRCIP 337
QY 228 -----LGIHLMD 233
DB 338 RMDWTMLGIEYD 349

```

RESULT 6

US-09-252-991A-17372

; Sequence 17372; Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 17372

; LENGTH: 401

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; US-09-252-991A-17372

Query Match 21.0%; Score 291.5; DB 4; Length 401;

Best Local Similarity 35.4%; Pred. No. 1.6e-26;

Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;

```

QY 1 MEAAATLIRLTNK-VPEPFLVLLISGGHCLALVQVSDPFLLGKSLDIAFGDMLDKVA 59
DB 172 MEGHLLAPLADDPSPFPVCLVSGGHTLVADGCVYQILGESIDDAVGEAFDXTAK 231
QY 60 RLSLIKHPECSTMSGKAIHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTVDKIMK 119
DB 232 LIGL-GYP-----GGPEIRALARGTRGRVFRPRMTDRGLDPSFGGLKTFILN-TWQ 283

```

```

QY 120 KEKEBGEIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKQDLPONNAVVASG 179
DB 284 REVEAGDDEQ---TRCDIALAFQTAVENTLTIKCRAL---KQTEL--KN---LVYAG 332
QY 180 VASNFYIRRALIEILTNATQCTLLCPPLCTDNGIMIANNGIBRLRAGLGIHLDEGIRY 239
DB 333 VSAHQALRSGLKEMKGMKGQVFYARPRCTDNGAMIAVAGCQRLIAG---QHDGPAISV 389
QY 240 EPEKCP 245
DB 390 QPRWPM 395

```

RESULT 7

US-09-543-681A-6513

; Sequence 6513; Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543.681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 6513

; LENGTH: 357

; TYPE: PRT

; ORGANISM: Proteus mirabilis

; US-09-543-681A-6513

Query Match 20.3%; Score 280.5; DB 4; Length 357;

Best Local Similarity 32.7%; Pred. No. 3e-25;

Matches 81; Conservative 35; Mismatches 101; Indels 31; Gaps 8;

```

QY 1 MEAAATLIRLTNKV-EPPFLVLLISGGHCLALVQVSDPFLLGKSLDIAFGDMLDKVA 59
DB 129 MEGHLLAPLADDPSPFPVCLVSGGHTLVADGCVYQILGESIDDAVGEAFDXTAK 188
QY 60 RLSLIKHPECSTMSGKAIHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTVDKIMK 119
DB 189 LIGL-DYP-----GGPVLSKVAQGVBEGRFVFRPRMTDRGLDPSFGGLKTFIANTIQ 241
QY 120 KEKEBGEIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQKQDLPONN-AVLVASG 178
DB 242 NDDSE-----QTRADIAAFEDAVVDTLAIKCRRA-----LEQYGRKRLVMAG 284
QY 179 GVASNFYIRRALIEILTNATQCTLLCPPLCTDNGIMIANNGIBRLRAGL-GIHLDEBI 237
DB 285 GVSANRTLRKAKAMIMEQVGEVYARPRCTDNGAMIAVAGIRKGGTEGPL---GV 340
QY 238 RYEPKCP 245
DB 341 TVPRRPM 348

```

RESULT 8

US-08-087-797-3

; Sequence 3; Application US/08087797

; Patent No. 5543312

; GENERAL INFORMATION:

; APPLICANT: Mellors, Alan

; APPLICANT: Lo, Reggie Y.C.

; APPLICANT: Abdullah, Khalid M.

; TITLE OF INVENTION: Pasteurella Haemolytica

; TITLE OF INVENTION: Glycoproteinase

; TITLE OF INVENTION: Gene and the Purified Enzyme

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell, Selczer, Park & Gibson, P.A.

; STREET: 1211 East Morehead Street,

CITY: Charlotte
STATE: No. 5543312th Carolina
COUNTRY: United States
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: #1.0, Version #1.25
APPLICATION NUMBER: US/08/087,797
FILING DATE: 14-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Layton, Jr., Samuel G.
REGISTRATION NUMBER: 22807
REFERENCE/DOCKET NUMBER: 3374-80
TELEPHONE: 704 377 1561
TELEFAX: 704 334 2014
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-087-797-3

Query Match 20.1%; Score 279; DB 1; Length 342;
Best Local Similarity 33.6%; Pred. No. 4.2e-25;
Matches 83; Conservative 32; Mismatches 108; Indels 24; Gaps 8;

QY 1 MEAAHATIRLT-TNKVEPPLVLLISGGHCLALVGVSDPFLIGSLDIAFGMDLKVAR 59
DB 112 MEGHIAEMLENDNPEPFVALVSGHQTOLISVIGIYGYSIDDAAGAEFDKAK 171
QY 60 RLSLIKHPECSTMSGGKAIEHLAKQGNRPHFDIKPPLHAKNCDPFTGLQHTYDKI 119
DB 172 LLGL-DYP-----GGPLSLKVAAGCTAGRFVFPPTMDRPGIDPSFSLKTFPAANTIRD 224
QY 120 KEKEGIEKQILSSADIAATVQHTMACHLVKTTRAILFCQKRDLPQNN-AVLVAGS 178
DB 225 NXXXXGXTHDDQ---TRADIAAFEDAVVDLTMIKKRA-----LDQTFKRLVWAG 272
QY 179 GVASNFYIRALLETITNAQCTLLCPPLCTDNGIMIANWGIERLAGIGILDIEGIR 238
DB 273 GVSANRTLRAXLAEMWKKRGEVIFYARPEFTDNGAMITAVGAVRFA--GATADL-GVS 329
QY 239 YEPKCP 245
DB 330 VRPRWPL 336

RESULT 9
US-10-067-443-4
Sequence 4; Application US/10067443
Patent No. 6642041
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
TITLE OF INVENTION: SPINAL CORD, MP-1
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 421

TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-067-443-4

Query Match 19.9%; Score 276; DB 4; Length 421;
Best Local Similarity 30.4%; Pred. No. 1.4e-24;
Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

QY 1 MEAAHATIRLT-TNKVEPPLVLLISGGHCLALVGVSDPFLIGSLDIAFGMDLKVAR 59
DB 134 MRAHASTILLVDVAFPPSAVLISGHALISVAEVEKFKYQGSVSGSPGICIKVAR 193
QY 60 RLSLIKHPECSTMSGGKAIEHLAKQGN---RPHFDIKPPLHAKNCDPFTGLQHTYDKI 116
DB 194 QUGDL-GSEFDGIVHGAAYEILASRASADGHLRYPLFNVPRKANNPQIKSYTLNLE 252
QY 117 IMKEKEGIEKQILSSADIAATVQHTMACHLVKTTRAILFCQKRDLPQNNALVA 176
DB 253 RLKNSETSID-----IPDFCASIQNTVAHISSKLHIFPESLSEQETLPKQ---LVI 302
QY 177 GGVASNFYIRALLETITNAQCTLLCPPLCTDNGIMIANWGIERLAGIGILDIEG 236
DB 303 GGVANAOYIFGALISKLSAAHVVTTIKVLLSLCTDNAEMIAVSGL-----LMLVNRSEA 356
QY 237 IRYEP-----KCPLGVDISKVEGASIKVPLQKM 265
DB 357 IMWRPNIDPTIYAHARSDIGTDSSET---IDTPRRKL 392

RESULT 10
US-10-067-443-28
Sequence 28; Application US/10067443
Patent No. 6642041

GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
TITLE OF INVENTION: SPINAL CORD, MP-1
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 28
LENGTH: 421
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-067-443-28

Query Match 19.9%; Score 276; DB 4; Length 421;
Best Local Similarity 30.4%; Pred. No. 1.4e-24;
Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

QY 1 MEAAHATIRLT-TNKVEPPLVLLISGGHCLALVGVSDPFLIGSLDIAFGMDLKVAR 59
DB 134 MRAHASTILLVDVAFPPSAVLISGHALISVAEVEKFKYQGSVSGSPGICIKVAR 193
QY 60 RLSLIKHPECSTMSGGKAIEHLAKQGN---RPHFDIKPPLHAKNCDPFTGLQHTYDKI 116
DB 194 QUGDL-GSEFDGIVHGAAYEILASRASADGHLRYPLFNVPRKANNPQIKSYTLNLE 252
QY 117 IMKEKEGIEKQILSSADIAATVQHTMACHLVKTTRAILFCQKRDLPQNNALVA 176
DB 253 RLKNSETSID-----IPDFCASIQNTVAHISSKLHIFPESLSEQETLPKQ---LVI 302
QY 177 GGVASNFYIRALLETITNAQCTLLCPPLCTDNGIMIANWGIERLAGIGILDIEG 236
DB 303 GGVANAOYIFGALISKLSAAHVVTTIKVLLSLCTDNAEMIAVSGL-----LMLVNRSEA 356
QY 237 IRYEP-----KCPLGVDISKVEGASIKVPLQKM 265

Db 357 IMMPPNDIPDTIYAHARSDIGTDASSEI-----IDTPRRRL 392

RESULT 11

US-08-087-797-2

; Sequence 2, Application US/08087797

; Patent No. 5543312

; GENERAL INFORMATION:

; APPLICANT: Mellores, Alan

; APPLICANT: Lo, Reggie Y.C.

; APPLICANT: Abdullah, Khalid M.

; TITLE OF INVENTION: Pasteurella Haemolytica

; TITLE OF INVENTION: Glycoprotease

; TITLE OF INVENTION: Gene and the Purified Enzyme

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.

; STREET: 1211 East Morehead Street,

; CITY: Charlotte

; STATE: No. 5543312th Carolina

; COUNTRY: United States

; ZIP: 28234

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/087,797

; FILING DATE: 14-JUL-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Layton, Jr., Samuel G.

; REGISTRATION NUMBER: 22807

; REFERENCE/DOCKET NUMBER: 3374-80

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 704 377 1561

; TELEFAX: 704 334 2014

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 325 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-087-797-2

Query Match 19.5%; Score 270.5; DB 1; Length 325;
Best Local Similarity 35.0%; Pred. No. 4.2e-24;

Matches 79; Conservative 26; Mismatches 102; Indels 19; Gaps 6;

QY 1 MEAHALITRL-TNKVEPPLVLLISGHCGLALVGVSDPFLIGKSLDIPAGMDLKVAR 59
DB 112 MEGLHAPMLIEDNPPAPFFVALLVSGGHTQLISVTIGQELGESIDDAAGAFKTK 171
QY 60 RLSLIKHPECSTMSGKAIHLAKOGRFHPDIKPLHAKKCDFFSTGLQHTVDKIMK 119
DB 172 LGLG-DYF-----AGVAMSKLASBSTPNRFRPMPMDRDLGDFSGGLKTFPAANTIK 224
QY 120 KEKEGIEKGQILSSADIAATVQHTMACHLVKTHTRAILFCOKRDLDPONNAVIVASGG 179
DB 225 NLNENGLDEO--TKCDIAHAFQAV-----VDTILIKCK-RALSQYGYKRLVMAGG 273
QY 180 VASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIMANGIERLR 225
DB 274 VSANKQIRADLAEMMKLKGIVFPFRPQCTDNGMIAVYGTFLRK 319

RESULT 12

US-09-489-039A-9221

; Sequence 9221, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 9221

; LENGTH: 343

; TYPE: PR

; ORGANISM: Klebsiella pneumoniae

; US-09-489-039A-9221

Query Match 19.4%; Score 268.5; DB 4; Length 343;
Best Local Similarity 32.7%; Pred. No. 8.1e-24;
Matches 81; Conservative 31; Mismatches 105; Indels 31; Gaps 8;

QY 1 MEAHALITRL-TNKVEPPLVLLISGHCGLALVGVSDPFLIGKSLDIPAGMDLKVAR 59
DB 118 MEGLHAPMLIEDNPPAPFFVALLVSGGHTQLISVTIGQELGESIDDAAGAFKTK 177
QY 60 RLSLIKHPECSTMSGKAIHLAKOGRFHPDIKPLHAKKCDFFSTGLQHTVDKIMK 119
DB 178 LGLG-DYF-----GGPMLSKMASQCTBGRFVPRPMDRDLGDFSGGLKTFPAANTIRS 230
QY 120 KEKEGIEKGQILSSADIAATVQHTMACHLVKTHTRAILFCOKRDLDPONN-AVIVASG 178
DB 231 NGDSE-----QTRADIAFAFEDAVVDTLMIKCRR-----LQTFKRLVMAG 273
QY 179 GVSANFYIRALEILTNATQCTLLCPPRLCTDNGIMIMANGIERLRAGILHIDE-GI 237
DB 274 GVSANFYIRALEILTNATQCTLLCPPRLCTDNGIMIMANGIERLRAGILHIDE-GI 237
QY 238 RYPCKPL 245
DB 330 TYRPRMPL 337

RESULT 13

US-09-107-532A-6609

; Sequence 6609, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: Jul 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Denke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6609:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1..363
SEQUENCE DESCRIPTION: SEQ ID NO: 6609:
US-09-107-552A-6609

Query Match 18.8%; Score 261; DB 4; Length 363;
Best Local Similarity 33.8%; Pred. No. 7, 3e-23;
Matches 75; Conservative 37; Mismatches 78; Indels 32; Gaps 9;

QY 1 MEAAHATIRLTNNKVEPPLVLLISGGHCILALVGVSDFLLGKSLDIAPGMDLVARR 60
DB 140 MAGHIVARLVKPFQPLMALIVSGHTELVYMGEDGSEYIIGETRDAAAGAYDKVGRV 199

QY 61 LSLIKPECTSMGSGKAIEHLAKOG-NRPHDIKPEPLHAKNCDFSTGLOHVTDKIIMK 119
DB 200 LGL-----SYPSKEIDQLAHQGRKNYH--PRAMIEDYDPSFSGLSKAPINLVHN 250

QY 120 -KEKEGIEKGQILSSADIATVQHTWACHLVKTRTHAILFCRKORDLLPQNNAV--LVA 176
DB 251 AQQGEEDLDK-----DLASFQASVYDVLTITLRA---C-----QYTPKQVLV 293

QY 177 SGVAASFYIRRALILTN--TQCTLLCPPRLCTDNGIMI 216
DB 294 AGVAANQGLREGLOALSAKLPELVLPRLCGDNAMMI 335

RESULT 14
US-08-987-121A-4
Sequence 4, Application US/08987121A
GENERAL INFORMATION:
APPLICANT: Hoskins, Jo Ann
APPLICANT: Tang, Joseph Chion-Chung
APPLICANT: Treadway, Patti Jean
TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
TITLE OF INVENTION: GCP
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: U.S.
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/987,121A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39,872
REFERENCE/DOCKET NUMBER: X-
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-987-121A-4

Query Match 18.3%; Score 254; DB 3; Length 336;
Best Local Similarity 32.4%; Pred. No. 4, 6e-22;
Matches 73; Conservative 36; Mismatches 86; Indels 30; Gaps 8;

QY 1 MEAAHATIRLTNNKVEPPLVLLISGGHCILALVGVSDFLLGKSLDIAPGMDLVARR 60
DB 115 MAGHIVARLVKPFQPLMALIVSGHTELVYMGEDGSEYIIGETRDAAAGAYDKVGRV 174

QY 61 LSLIKPECTSMGSGKAIEHLAKOG-NRPHDIKPEPLHAKNCDFSTGLOHVTDKIIMK 120
DB 175 MGL-----TYPAGEIDELAHQGHDI-YDPPRAIKEDNIEFSGLSAPINLVHNA 226

QY 121 EKEGIEKGQILSSADIATVQHTWACHLVKTRTHAILFCRKORDLLPQNNAVLVASGV 180
DB 227 E-----QKESLST-EDLCSFQAAVMDILMAKTKAL-----EKYPK--TLVYAGV 272

QY 181 ASNFYIRRALILTN--TQCTLLCPPRLCTDNGIMI-----WN 219
DB 273 AANKGLRRL--ATBITDVNVIIIPRLCGDNAMGIVAVSSEWN 315

RESULT 15
US-08-961-083-52
Sequence 52, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-961-083-52

Query Match 18.1%; Score 250; DB 3; Length 335;
Best Local Similarity 32.0%; Pred. No. 1, 4e-21;
Matches 72; Conservative 36; Mismatches 87; Indels 30; Gaps 8;

```

QY 1 MEAHATIRLTNKVPEPFLVLLISGHCILALVQVSDPFLGSGUDIPAGDMLDXVAR 60
Db 114 MAGHLMMAQSVPEPFLALALVSGHTELYVSEAGDYKIVGETRDAVGEAYDKGRV 173
QY 61 LSLIKHPECSTMSGKALEHLAKQGNRPHFDIKPPLHAKNCDFSTGLQHVTDKIMCK 120
Db 174 MGL-----TYPAGREIDBLAHQODI-YDEPRAMIKEDNLEFSPGSKSAFINLHNHNA 225
QY 121 EKEGIEKGQILLSADIAATVQHTMACHLVKTTRAILFCQORDLLPQNNAVLVASGV 180
Db 226 E-----OKGESIST-BDLCASFOAAVMDILMAKTKKAL-----EKYPVK--ILVVAGGV 271
QY 181 ASNFYIRALEILMNAQCTLLCPPRLCTDNGIMIA-----WN 219
Db 272 AANKGLRERL--AAETIDVKYIIPPLRLCGDNAGMIAIASVSXWN 314

```

Search completed: February 16, 2005, 13:10:25
 Job time : 16.962 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:08:47 ; Search time 154.686 Seconds
(without alignments)
563.995 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHLITRLTNKVEFPFLV.....DISKVGASIKVQLKMEI 267

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	267	14	US-10-067-443-22 Sequence 22, Appl
2	1385	100.0	267	15	US-10-649-273-22 Sequence 22, Appl
3	1385	100.0	267	15	US-10-651-722-22 Sequence 22, Appl
4	1385	100.0	414	14	US-10-067-443-22 Sequence 2, Appl1
5	1385	100.0	414	15	US-10-649-273-22 Sequence 2, Appl1
6	1385	100.0	414	15	US-10-651-722-22 Sequence 2, Appl1
7	1358.5	98.1	439	14	US-10-067-443-19 Sequence 19, Appl
8	1358.5	98.1	439	15	US-10-649-273-19 Sequence 19, Appl
9	1358.5	98.1	439	15	US-10-651-722-19 Sequence 19, Appl
10	1348	97.2	414	14	US-10-012-140-5 Sequence 5, Appl1
11	1105	79.8	364	15	US-10-094-749-2039 Sequence 2039, Ap
12	422.5	30.5	445	15	US-10-424-599-209259 Sequence 209259, Ap
13	409.5	29.6	463	14	US-10-067-443-3 Sequence 3, Appl1

14	409.5	29.6	463	15	US-10-649-273-3	Sequence 3, Appl1
15	409.5	29.6	444	15	US-10-651-722-3	Sequence 3, Appl1
16	359	25.9	446	15	US-10-437-963-113732	Sequence 113732, A
17	342	24.7	251	14	US-10-081-051-9	Sequence 9, Appl1
18	321.5	23.2	382	14	US-10-282-122A-50858	Sequence 50858, A
19	317	22.9	333	14	US-10-012-140-25	Sequence 25, Appl1
20	314.5	22.7	312	15	US-10-282-122A-44459	Sequence 44459, A
21	308.5	22.3	348	15	US-10-282-122A-63156	Sequence 63156, A
22	304.5	22.0	343	15	US-10-282-122A-67227	Sequence 67227, A
23	294	21.2	347	14	US-10-012-140-24	Sequence 24, Appl
24	291.5	21.0	341	9	US-09-815-242-11798	Sequence 11798, A
25	291.5	21.0	342	15	US-10-282-122A-66200	Sequence 66200, A
26	289.5	20.9	342	15	US-09-815-242-11043	Sequence 11043, A
27	289.5	20.9	342	15	US-10-282-122A-58204	Sequence 58204, A
28	280.5	20.3	340	15	US-10-282-122A-68438	Sequence 68438, A
29	279.5	20.2	337	9	US-09-815-242-13780	Sequence 13780, A
30	279.5	20.2	337	15	US-10-282-122A-75455	Sequence 75455, A
31	279.5	20.2	337	15	US-10-282-122A-78161	Sequence 78161, A
32	277.5	20.0	335	15	US-10-282-122A-55404	Sequence 55404, A
33	277	20.0	309	15	US-10-282-122A-61538	Sequence 61538, A
34	277	20.0	341	15	US-10-282-122A-67993	Sequence 67993, A
35	276	19.9	421	14	US-10-067-443-4	Sequence 4, Appl1
36	276	19.9	421	14	US-10-067-443-28	Sequence 28, Appl1
37	276	19.9	421	15	US-10-649-273-4	Sequence 4, Appl1
38	276	19.9	421	15	US-10-649-273-28	Sequence 28, Appl1
39	276	19.9	421	15	US-10-651-722-4	Sequence 4, Appl1
40	276	19.9	421	15	US-10-651-722-28	Sequence 28, Appl1
41	275.5	19.9	337	9	US-09-815-242-10304	Sequence 10304, A
42	275.5	19.9	337	15	US-10-282-122A-56695	Sequence 56695, A
43	263.5	19.0	354	15	US-10-282-122A-65768	Sequence 65768, A
44	263	19.0	341	15	US-10-282-122A-69613	Sequence 69613, A
45	257	18.6	338	15	US-10-282-122A-57817	Sequence 57817, A

ALIGNMENTS

RESULT 1
US-10-067-443-22 Application US/10067443
; Sequence 22, Appl1
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-22

Query Match 100.0%; Score 1385; DB 14; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHLITRLTNKVEFPFLVLLISGHCILALVGVSPDLILGKSLDIAFGMLDKVAR 60
DB 1 MEAHLITRLTNKVEFPFLVLLISGHCILALVGVSPDLILGKSLDIAFGMLDKVAR 60
QY 1 LSLIHGPCSTMSGGKALHAKQNRFFHFDIKPLHAKKNDPFTGLOHTYDKIIMKX 120
DB 61 LSLIHGPCSTMSGGKALHAKQNRFFHFDIKPLHAKKNDPFTGLOHTYDKIIMKX 120
QY 121 EKESGIEKGQILSSAADIATVQHTMACHLVKRTIRAILPCKQRDLIPONNAVVASGCV 180

Db 121 EKEGIEKQIISADIAATVQHTMACHLVYRTHRAILFCQORDLLPQNNAVLVAASGV 180
QY 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRYE 240
Db 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRYE 240
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 2
US-10-649-273-22
Sequence 22, Application US/10649273
Publication No. US20040043407A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
FILE REFERENCE: D0073 CNT
CURRENT APPLICATION NUMBER: US/10/649,273
PRIOR FILING DATE: 2003-08-27
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 22
LENGTH: 267
TYPE: PRT
ORGANISM: homo sapiens
US-10-649-273-22

Query Match 100.0%; Score 1385; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAAHATIRLTNKKVEPFPVLVLLISGHCILALVQGVSDFLILGKSIDIAFGMDLKVARR 60
Db 1 MEAAHATIRLTNKKVEPFPVLVLLISGHCILALVQGVSDFLILGKSIDIAFGMDLKVARR 60
QY 61 LSLIKHPECSTWGGKAIETHAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTVDKTIIMKK 120
Db 61 LSLIKHPECSTWGGKAIETHAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTVDKTIIMKK 120
QY 121 EKEGIEKQIISADIAATVQHTMACHLVYRTHRAILFCQORDLLPQNNAVLVAASGV 180
Db 121 EKEGIEKQIISADIAATVQHTMACHLVYRTHRAILFCQORDLLPQNNAVLVAASGV 180
QY 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRYE 240
Db 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRYE 240
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 3
US-10-651-722-22
Sequence 22, Application US/10651722
Publication No. US20040048302A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
FILE REFERENCE: D0073 DIV
CURRENT APPLICATION NUMBER: US/10/651,722
PRIOR FILING DATE: 2003-08-29
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 22
LENGTH: 267
TYPE: PRT
ORGANISM: homo sapiens
US-10-651-722-22

Query Match 100.0%; Score 1385; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAAHATIRLTNKKVEPFPVLVLLISGHCILALVQGVSDFLILGKSIDIAFGMDLKVARR 60
Db 1 MEAAHATIRLTNKKVEPFPVLVLLISGHCILALVQGVSDFLILGKSIDIAFGMDLKVARR 60
QY 61 LSLIKHPECSTWGGKAIETHAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTVDKTIIMKK 120
Db 61 LSLIKHPECSTWGGKAIETHAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTVDKTIIMKK 120
QY 121 EKEGIEKQIISADIAATVQHTMACHLVYRTHRAILFCQORDLLPQNNAVLVAASGV 180
Db 121 EKEGIEKQIISADIAATVQHTMACHLVYRTHRAILFCQORDLLPQNNAVLVAASGV 180
QY 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRYE 240
Db 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRYE 240
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 4
US-10-067-443-2
Sequence 2, Application US/10067443
Publication No. US20030082782A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 414
TYPE: PRT
ORGANISM: Homo sapiens
US-10-067-443-2

Query Match 100.0%; Score 1385; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAAHATIRLTNKKVEPFPVLVLLISGHCILALVQGVSDFLILGKSIDIAFGMDLKVARR 60
Db 148 MEAAHATIRLTNKKVEPFPVLVLLISGHCILALVQGVSDFLILGKSIDIAFGMDLKVARR 207
QY 61 LSLIKHPECSTWGGKAIETHAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTVDKTIIMKK 120
Db 208 LSLIKHPECSTWGGKAIETHAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTVDKTIIMKK 267
QY 121 EKEGIEKQIISADIAATVQHTMACHLVYRTHRAILFCQORDLLPQNNAVLVAASGV 180

```
Db      268 EKEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLIPONNAVIVASGV 327
Qy      181 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 240
Db      328 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 387
Qy      241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db      388 PKCPLGVDISKEVGEASIKVPOLKMEI 414
```

```
RESULT 5
US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2
```

```
Query Match      100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MEAAHATIRLTNKKVEFPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVARR 60
Db      148 MEAAHATIRLTNKKVEFPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVARR 207
Qy      61 LSLIKHPECGSTMSGGKAI EHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHYTDKIMKK 120
Db      208 LSLIKHPECGSTMSGGKAI EHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHYTDKIMKK 267
Qy      121 EKEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLIPONNAVIVASGV 180
Db      268 EKEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLIPONNAVIVASGV 327
Qy      181 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 240
Db      328 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 387
Qy      241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db      388 PKCPLGVDISKEVGEASIKVPOLKMEI 414
```

```
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2
```

```
Query Match      100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MEAAHATIRLTNKKVEFPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVARR 60
Db      148 MEAAHATIRLTNKKVEFPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVARR 207
Qy      61 LSLIKHPECGSTMSGGKAI EHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHYTDKIMKK 120
Db      208 LSLIKHPECGSTMSGGKAI EHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHYTDKIMKK 267
Qy      121 EKEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLIPONNAVIVASGV 180
Db      268 EKEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLIPONNAVIVASGV 327
Qy      181 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 240
Db      328 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 387
Qy      241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db      388 PKCPLGVDISKEVGEASIKVPOLKMEI 414
```

```
RESULT 7
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19
```

```
Query Match      98.1%; Score 1358.5; DB 14; Length 439;
Best Local Similarity 91.1%; Pred. No. 2.5e-136;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```

```
Qy      1 MEAAHATIRLTNKKVEFPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVARR 60
Db      148 MEAAHATIRLTNKKVEFPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVARR 207
Qy      61 LSLIKHPECGSTMSGGKAI EHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHYTDKIMKK 120
Db      208 LSLIKHPECGSTMSGGKAI EHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHYTDKIMKK 267
Qy      121 EKEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLIPONNAVIVASGV 180
Db      268 EKEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLIPONNAVIVASGV 327
Qy      181 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 240
Db      328 ASNFYIRRALIELTNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGILHIDIGIRYE 387
Qy      241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db      388 PKCPLGVDISKEVGEASIKVPOLKMEI 414
```

```

Db      268 EKEGGIFLISKVEQINIPGLCKIAHPCRYEKGQILSSADIAATVQHTMACHLVKRTH 327
      156 RAIFCCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPRLCTDNGIM 215
      328 RAIFCCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPRLCTDNGIM 387
Qy      216 IAWNGIERLRAGLGILHDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 267
      388 IAWNGIERLRAGLGILHDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 439

```

RESULT 8

```

US-10-649-273-19
; Sequence 19, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

```

Query Match 98.1%; Score 1358.5; DB 15; Length 439;

Best Local Similarity 91.1%; Pred. No. 2.5e-136; Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

Qy      1 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDPFLIGKSLDIAPGMDLVKVAR 60
      148 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDPFLIGKSLDIAPGMDLVKVAR 207
      61 LSLIHPECSTMSGGKALIEHLAKQGNRPFDIKPPLHAKNCDPFTGLQHTVDKTIIMKK 120
      208 LSLIHPECSTMSGGKALIEHLAKQGNRPFDIKPPLHAKNCDPFTGLQHTVDKTIIMKK 267
Qy      121 EKEGGI-----EKGQILSSADIAATVQHTMACHLVKRTH 155
      268 EKEGGIFLISKVEQINIPGLCKIAHPCRYEKGQILSSADIAATVQHTMACHLVKRTH 327
Qy      156 RAIFCCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPRLCTDNGIM 215
      328 RAIFCCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPRLCTDNGIM 387
Qy      216 IAWNGIERLRAGLGILHDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 267
      388 IAWNGIERLRAGLGILHDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 439

```

RESULT 9

```

US-10-651-722-19
; Sequence 19, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05

```

```

; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

```

Query Match 98.1%; Score 1358.5; DB 15; Length 439;

Best Local Similarity 91.1%; Pred. No. 2.5e-136; Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

Qy      1 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDPFLIGKSLDIAPGMDLVKVAR 60
      148 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDPFLIGKSLDIAPGMDLVKVAR 207
      61 LSLIHPECSTMSGGKALIEHLAKQGNRPFDIKPPLHAKNCDPFTGLQHTVDKTIIMKK 120
      208 LSLIHPECSTMSGGKALIEHLAKQGNRPFDIKPPLHAKNCDPFTGLQHTVDKTIIMKK 267
Qy      121 EKEGGI-----EKGQILSSADIAATVQHTMACHLVKRTH 155
      268 EKEGGIFLISKVEQINIPGLCKIAHPCRYEKGQILSSADIAATVQHTMACHLVKRTH 327
Qy      156 RAIFCCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPRLCTDNGIM 215
      328 RAIFCCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPRLCTDNGIM 387
Qy      216 IAWNGIERLRAGLGILHDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 267
      388 IAWNGIERLRAGLGILHDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 439

```

RESULT 10

```

US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leibniz, Kevin R.
; APPLICANT: Kappeler-Liebermann, Rosana
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

```

Query Match 97.3%; Score 1348; DB 14; Length 414;

Best Local Similarity 97.4%; Pred. No. 3.1e-135; Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

Qy      1 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDPFLIGKSLDIAPGMDLVKVAR 60
      148 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDPFLIGKSLDIAPGMDLVKVAR 207

```

QY	61	LSLHGHPCSTMSGGKALEHIAKQNRFFEDJKPPLAHAKNCDPSTGLOHTVDKIMKK	120
Db	208	LSLHKPCSTMSGGKALEHIAKQNRFFEDJKPPLAHAKNCDPSTGLOHTVDKNNNR	267
QY	121	EKEGIEKQILLSSADIAATVQHTMACHLVKTRTHAILFCQRDILLPONNAVLAASGV	180
Db	268	KQEGIEKQILLSSADIAATVQHTMACHLVKTRTHAILFCQRDILLPONNAVLAASGV	327
QY	181	ASNPIRRALEILTNAOTCTLLCPPEPLCTDNGIMIANNGIERLRGLGIHIDIGIRIR	240
Db	328	ASNPIRRALEILTNAOTCTLLCPPEPLCTDNGIMIANNGIERLRGLGIHIDIGIRIR	387
QY	241	PKCPAGVDISKVEGASIKVPOLKKEI	267
Db	388	PKCPAGVDISKVEGASIKVPOLKKEI	414

```

RESULT 11
US-10-094-749-2039
? Sequence 2039, Application US/10094749
? Publication No. US20030219741A1
? GENERAL INFORMATION:
? APPLICANT: ISOGAI, TAKAO
? APPLICANT: SUGIYAMA, TOMOYASU
? APPLICANT: OTSUKI, TETSUJI
? APPLICANT: WAKAMATSU, AI
? APPLICANT: SATO, HIROYUKI
? APPLICANT: ISHII, SHIZUKU
? APPLICANT: YAMAMOTO, JUN-ICHI
? APPLICANT: ISONO, YUUKO
? APPLICANT: HIO, YURI
? APPLICANT: OTSUKA, KAORU
? APPLICANT: NAGAI, KEIICHI
? APPLICANT: IRIE, RYOTARO
? APPLICANT: TAKECHIKA, ICHIRO
? APPLICANT: SEKI, NAOHICO
? APPLICANT: YOSHIKAWA, TUTORU
? APPLICANT: OTSUKA, MOTOTYUKI
? APPLICANT: NAGAHARI, KENJI
? APPLICANT: MASUHO, YASUHIKO
? TITLE OF INVENTION: NOVEL FULF.-LENGTH CDNA
? FILE REFERENCE: 084335/0160
? CURRENT APPLICATION NUMBER: US/10/094,749
? CURRENT FILING DATE: 2002-03-12
? PRIOR APPLICATION NUMBER: 60/350,435
? PRIOR FILING DATE: 2002-01-24
? PRIOR APPLICATION NUMBER: JP 2001-328381
? PRIOR FILING DATE: 2001-09-14
? NUMBER OF SEQ ID NOS: 3361
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 2039
? LENGTH: 364
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-094-749-2039

```

Query Match	79.8%	Score 1105	DB 15	length 364
Best Local Similarity	98.6%	Pred. No. 2.6e-109		
Matches 214	Conservative 0	Mismatches 3	Indels 0	Gaps 0
QY	1	MEAAATLTRLTNKVEFPFLVLLISGGHCLLALVQGSDFLLGKSIDIAFGMDLDRVAR	60	
Db	148	MEAAATLTRLTNKVEFPFLVLLISGGHCLLALVQGSDFLLGKSIDIAFGMDLDRVAR	207	
QY	61	LSLTKHPGCSMTSGSKAEHLAKQNRPHPIKPLHLAKNODPFTGQHTDGLIMKK	120	
Db	208	LPLKHPGCSMTSGSKAEHLAKQNRPHPIKPLHLAKNODPFTGQHTDGLIMKK	267	
QY	121	EKEGGIEKQILLSSAADIAATVQHTMACHLYKRTIRALIFCKQRIILLPQNNANVLVASGCV	180	
Db	268	EKEGGIEKQILLSSAADIAATVQHTMACHLYKRTIRALIFCKQRIILLPQNNANVLVASGCV	327	
QY	181	ASNFYIRRALEILTNATQCTLLCPEPRLCTDNGIMIA	217	

Db 328 ASNFCIRALEILTNAQTCLCPPPRLCTDNGIMIA 364

```

US-10-424-599-209259
RESULT 12
US-10-424-599-209259
Sequence 209259, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ. ID NOS: 285684
SEQ ID NO 209259
LENGTH: 445
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_3098C.1.pep
US-10-424-599-209259

```

	Query Match	30.5%	Score 422.5;	DB 15;	Length 445;
	Best Local Similarity	41.6%;	Pred. No. 4.1e-36;		
	Matches 104;	Conservative 32;	Mismatches 89;	Indels 25;	Gaps 6
Qy	1 MEAAHLIRLTNNK--VEEPPFLVLISGGCHLALVQGVSDPFLGKSLDIAGDMLDKVAR	59			
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
Dd	165 MEAAHVALGRLLKEQLQFFPMALLISSGNMLLVLRDLQGYIQLGTITDADIGEAIDTKAR	224			
Qy	60 RLSTLIKHPECSWTMSGKAIEHLAKQGNRFHEIDIKRPLHAANCDSEFTGLQ----	HVTD 114			
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
Dd	225 WLGIDLK----RSGGPAIEKLAMEGNABSVKFSIPMKQHDCNFSYAGLKTVQVRLAIES	279			
Qy	115 KIIMKKERBEETEGQILSSAADIAATVQHTMAChLYKTRTALLFCQRDILLPONNAV	174			
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
Dd	280 KKIDAKPISSASNGDRLLSRADIINASFORAVAILBERCERALOWALKMEPSIIS--L	335			
Qy	175 VASGSVASNPFIIRBALRIITNAOTCTLCPPRPRLCTONGIMIANGIERTLAGGILHDI	234			
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
Dd	336 VWSGSVASNQIVRARLDMVKRNGLQLVCPPRLCTDNGVMIAWTEIHFRMG-----	388			
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
Qy	235 EGIRYEPKCP 244				
	. .				
Dd	389 ---RYDEPPP 395				

```

RESULT 13
US-10-067-443-3
; Sequence 3, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```


US-10-067-443-3

Query Match	29.6%;	Score 409.5;	DB 14;	Length 463;
Best Local Similarity	36.6%;	Pred. No. 1.1e-34;		
Matches 100; Conservative	37;	Mismatches 91;	Indels 45;	Gaps 6

```

Qy 1 MEAHLTRLT-NKKEPFLVLLIGGSHGLAVQSVSPFLIGKSLDAPGMLRVAR 59
Db 192 MEAHLVARLVQEGELSPFMALLIGGNHVLVLANHGQYTGTLTVDDALGAPDKTAK 251
Qy 60 RLSLIKPECSMGKALIEHLAKGNRFHEDIKPLLHAKNCDSFSTGLQHVTDKIIMK 119
Db 252 WLGLDMH-----RSGGPAVEEIALTGDAKSVLENVPMKHCNCFSYAGIKTVRLAIEA 306
Qy 120 KEKEGIEIKGQLISAADIATVQHTMAQLVAKTRTALIFCKORBLPQNNAVLASGG 179
Db 307 KE-----IRRADIASFQVAVVLAHLEKCEKRALIDMALE---LEPSIGMAYISGG 353
Qy 180 VASNFYIRALEILTNATQCTLLCPRPRLCTDNGIMIAVNGIERLAPAGLGIIMDEGIRY 239
Db 354 VASNFYVRLANNIVENKRLKIVCPRPRLCTDNGVAVVATGLENHFRVG-----RY 403
Qy 240 E-----PKCPRLGVLSKEVGEA 256
Db 404 DPPRPATPEEDYVDLRPRWFLGEEYAKGRSEA 436

```

RESULT 14
US-10-649-273-3

Query Match	29.6%	Score 409.5	DB 15	Length 463
Best Local Similarity	36.6%	Pred. No. 1.1e-34		
Matches 100, Conservative	37	Mismatches 91	Indels 45	Gaps 6

```

Qy      1 MEAALATIRLT-NKKEPFPVLVLLIGGHCILALVOCSPFLLIGSLDIAPGMLERVAR 59
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      192 MEAAALVARLVEQELSPFPMALLISGGHNLVLIAHNGGQYQLGTTVDDALIGAPDKTAK 251
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      60 RLSLIKHEPCSTMSGKALIEHLAKONGNRHPIKRPYLHAKNCDSPTGLQJQHTYDKIMK 119
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      252 WLGIDMH-----RSGGPAAVEELATLGDGAKSVKPNVPMKHKONFVYAGLKTQVRLALIEA 306
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      120 KEKEGIEKGQILSSADIAATVQHTMACHLVKTRTHRALIFCKQORDLPONNAVLVASGG 179
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      307 KE-----IRNRADIAASFQVAVLHLEBEKERRALDWMLE---LEPSIKHWYISGG 353
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      180 VASNFYIRALEILTNATQCTLCPPRPLCTNGMIANNGAERLPAAGLHDIIGIRY 239
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      354 VASNKYVYLRLLNMYIENKNLKLVCPPPSLCTCTNGVNAVTLGLEHFRVG-----RY 403
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      240 E-----EKCPILGVDSIKVEGEA 256
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 404 DPPPATEPDYVDLRPRWPLGEEYAKGRSEA 436

```

RESULT 15
US-10-651-722-3
; Sequence 3, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-651-722-3

```

Query Match	29.6%	Score 409.5	DB 15	length 463
Best Match Similarity	36.6%	Pred. No. 1.1e-34		
Matches 100	37	Mismatches 91	Indels 45	Gaps 6

QY MEHAAITIRLT-NKVEPPLVYLLISGGCHLALVGVSDPILLSKSIDPAGDMIDKVAR 59
 Db MEHAAITIRLT-NKVEPPLVYLLISGGCHLALVGVSDPILLSKSIDPAGDMIDKVAR 59
 192 MEHAAITIRLT-NKVEPPLVYLLISGGCHLALVGVSDPILLSKSIDPAGDMIDKVAR 251
 QY RUSLIKHEPCSTMSGGKAEHLAKOGNRPHPDIKPEPLHIANCDPSTGLQHTDTKIMK 119
 Db RUSLIKHEPCSTMSGGKAEHLAKOGNRPHPDIKPEPLHIANCDPSTGLQHTDTKIMK 119
 252 WGLIDMH-----RSGGPAVEALTEALBGDAKSVFNPMPKXHCDCNFSYAGLTQVRLAIEA 306
 QY KEKEBEGIEKGQILISSADIAATVORTMACHVKRTHRIILFCORIDLPPONNAVLVASGG 179
 Db KEKEBEGIEKGQILISSADIAATVORTMACHVKRTHRIILFCORIDLPPONNAVLVASGG 179
 307 KE-----IRNRADIASFQVAVLHLEKCKERAILDALE---LEPISKHNVISGG 355
 QY VASNPFIYRALIEILTNAPQCTLLCEPPLRLCTDNGIMIANANGIEELRGLGIIDHIEGIRY 239
 Db VASNPFIYRALIEILTNAPQCTLLCEPPLRLCTDNGIMIANANGIEELRGLGIIDHIEGIRY 239
 354 VASNPFIYRALIEILTNAPQCTLLCEPPLRLCTDNGIMIANANGIEELRGLGIIDHIEGIRY 403
 QY E-----PKCPGVDSIKVEGEA 256
 Db E-----PKCPGVDSIKVEGEA 256
 404 DPEPPATBEDYVYDLRPRMPLGEEYVAAKGRSEA 436

Search completed: February 16, 2005, 13:28:30
Job time : 155.686 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:10:32 ; Search time 4248.78 Seconds
(without alignments)
3045.000 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALIRLRITNKVEPPFLV.....DISKEVGASIKVPQLKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1h
-O=/cg2_1/USPTO.spool/US10649273/runat_14022005_114702_16399/app_query.fasta_1.1429
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOBEFT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=humand0.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10649273.qcgn_1_1_8655.qrunat_14022005_114702_16399 -MCPU=6 -ICPU=3
-NO_MMAP -LARGOUDERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenEmbl.*

1: gb ba:*
2: gb hcg:*
3: gb in:*
4: gb om:*
5: gb ov:*
6: gb pat:*
7: gb ph:*
8: gb pl:*
9: gb pr:*
10: gb ro:*
11: gb sts:*
12: gb sy:*
13: gb un:*
14: gb vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	1416	6	AR541929 Sequence
2	1385	100.0	1526	6	AR428809 Sequence
3	1385	100.0	1908	9	BC011904 Homo sapi
4	1385	100.0	2197	6	AR428803 Sequence

5	1358.5	98.1	1387	6	AR428808	AR428808 Sequence
6	1358.5	98.1	1387	9	HSAA295148	AJ295148 Homo sapi
7	1348	97.3	1245	6	AX664697	AX664697 Sequence
8	1348	97.3	1820	6	AX664695	AX664695 Sequence
9	1231	88.9	1844	10	BC058172	BC058172 Mus muscu
10	1224	88.4	1017	10	BC038910	BC038910 Mus muscu
11	1204	86.9	2208	6	AX713716	AX713716 Sequence
12	1204	86.9	2208	6	AX055441	AX055441 Homo sapi
13	1116	80.6	1546	10	BC078974	BC078974 Rattus no
14	950.5	68.6	1109	10	BC051211	BC051211 Mus muscu
15	939	67.8	1522	5	BX934991	BX934991 Gallus ga
16	938	67.7	860	5	BX930694	BX930694 Gallus ga
17	934	67.4	1558	5	BX930963	BX930963 Gallus ga
18	831.5	60.0	84115	9	AC013468	AC013468 Homo sapi
19	824.5	59.5	14364	6	AR428807	AR428807 Sequence
20	666	48.1	249601	2	AC114153	AC114153 Rattus no
21	666	48.1	308652	2	AC121478	AC121478 Rattus no
22	652.5	47.1	256751	10	AC122925	AC122925 Mus muscu
23	610	44.0	875	6	CQ721898	CQ721898 Sequence
24	488.5	35.3	1474	3	AK113378	AK113378 Clona int
25	477.5	34.5	121251	5	AL591593	AL591593 Zebrafish
26	435	31.4	117322	5	AL672217	AL672217 Zebrafish
27	415.5	30.0	1576	3	AY051882	AY051882 Drosophi
28	415.5	30.0	1601	6	CQ606432	CQ606432 Sequence
29	415.5	30.0	3656	6	CQ606431	CQ606431 Sequence
30	415.5	30.0	14679	2	AC018262	AC018262 Drosophi
31	415.5	30.0	180263	3	AC010671	AC010671 Drosophi
32	415.5	30.0	207432	3	AE003513	AE003513 Drosophi
33	408	29.5	1672	8	AK070912	AK070912 Oryza sat
34	407.5	29.4	1443	8	AY024338	AY024338 Arabidops
35	407.5	29.4	1474	8	AY117283	AY117283 Arabidops
36	407.5	29.4	1567	8	AY063864	AY063864 Arabidops
37	401.5	29.0	1557	8	AY084577	AY084577 Arabidops
38	389	28.1	110000	2	AP006501	AP006501 Contination (9 of
39	369	26.6	260271	1	AE017258	AE017258 Molbachia
40	356.5	25.7	333800	1	SME591792	AL591792 Sinorhizo
41	356	25.7	20021	10	AF36796781	AF367967 Mus muscu
42	356	25.7	179252	10	AF131205	AF131205 Mus muscu
43	352	25.4	349116	1	AP003003	AP003003 Mesothizo
44	349	25.2	110000	1	AE017197	Continuation (2 of
45	342	24.7	756	6	AX685201	AX685201 Sequence

ALIGNMENTS

RESULT 1	AR541929	Sequence 177 from patent US 6743619.	1416 bp	DNA	linear	PAT 08-OCT-2004
LOCUS	AR541929					
DEFINITION	Sequence 177 from patent US 6743619.					
ACCESSION	AR541929					
VERSION	AR541929.1	GI:53934009				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1416)					
AUTHORS	Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F., Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R., Wang, D. and Drmanac, R.T.					
TITLE	Nucleic acids and polypeptides					
JOURNAL	Patent: US 6743619-A 177 01-JUN-2004;					
FEATURES	Location/Qualifiers					
source	1..1416					
ORIGIN	/organism="unknown"					
	/mol_type="genomic DNA"					

Alignment Scores:
Pred. No.: 1,37e-126
Score: 1385.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Length: 1416
Matches: 267
Conservative: 0
Mismatch: 0
Indels: 0

DB:	6	Gaps:	0
US-10-649-273-2_COPY_148_414 (1-267) x AR541929 (1-1416)			
Oy	1	MetGlu1aH1sAlaLeuThr11eArgLeuThrAsnLysValGluPheProPheLeuVal	20
Db	502	ATGAGGCTCATGCACCTTACTATTGGTTGACCAATTAAAGTAATTTCTTTTTCATT	561
Oy	21	LeuLeu11eSerG1yG1yH1sCysLeuLeu11aLeuValG1ng1yVal1SerAspPheLeu	40
Db	562	CTTTTGATTTCTGGAGGCTCATGCTGTGTGGCATTTAGTTCAAGAGTTTGAGATTTTCG	621
Oy	41	LeuLeuG1yLysSerLeuAsp11eAla1ProG1yAspMetLeuAspLysVal1AlaArg	60
Db	622	CTTTTGGAAAGCTTTTGGACATAGCACACAGTGCACAGCTTGCACAGTGCACAGAGA	681
Oy	61	LeuSerLeu11eLysH1s1ProG1yCysSerThrMetSerG1yG1yVal11eG1uH1s	80
Db	682	CTTTCTTTTAA1TAAACATCCAGAGTGTCTCCACCATGATGTGTGGAAAGCCATTGAACAT	741
Oy	81	Leu11aLysG1ng1yAsnArgPheH1sPheAsp11eLysProProLeuH1sH1sAlaLys	100
Db	742	TTGGCCAAACAAGGAATATAGATTTCATTTTGACATCAAACTCCCTTGCATCATGCTAA	801
Oy	101	AsnCysAspPheSerPheThrG1yLeuG1nH1sVal1ThrAspLys11e11MetLysLys	120
Db	802	AATGTGATTTTCTTTTACTGACCTTCACACCTTACTGTATTAATATGAAAAG	861
Oy	121	GluLysG1ng1yG1y11eGluLysG1yG1n11eLeuSerSer11a1Asp11eAla1a	140
Db	862	GAAAAGAGGAAGGATGAGAAAGGGGCAAACTGTCTTCACGACGACATTTGCTGC	921
Oy	141	ThrValG1nH1sThrMetAlaCysH1sLeuVal1LysArgThrH1sArgAla11eLeuPhe	160
Db	922	ACAGTACGACACACATGCGATGTCATTTGTGAAAABAACATCCGCGTATTTCTGTT	981
Oy	161	CysLysG1nArgAspLeuLeuProG1nAsnAsnAlaVal1LeuVal11aSerG1yG1yVal	180
Db	982	TGTAAAGCAGAGACCTGTACTCTCAAAATATATGACATGCTGTGCTGTGGTGTCTC	1041
Oy	181	AlaSerAsnPheYr11eArgArgAlaLeuG1n11eLeuThrAsnAla11eThrG1nCysThr	200
Db	1042	GCAAGTAACTCTATATCCGACAGCTCTGGAAATTTTAAACAAAGCAACAGTGCAT	1101
Oy	201	LeuLeuCysProProProArgLeuLeuCysThrAspAsnG1y11eMet11eAla11eTrpAsnG1y	220
Db	1102	TTGTGTGTCTCTCTCCACGACATGATGATATATGCACTTATGATTCATGCAATGAAATGT	1161
Oy	221	11eGluArgLeuArgAlaG1yLeuG1y11eLeuH1sAsp11eG1ng1y11eArgYrGlu	240
Db	1162	ATTGAAAGACATACGCTGCTGGCGCTTTTATCATGACATAGAAAGCATCCGCTATGAA	1221
Oy	241	ProLysCysProLeuG1yVal1Asp11eSerLysG1yVal1G1yG1uAlaSer11eLysVal	260
Db	1222	CCAAAAGTCTCTTGGAGTACATATCAAAAGAAAGTGGAGAGACTTTCATATAAAGTA	1281
Oy	261	ProG1nLeuLysMetG1u11e267	
Db	1282	CCACATTAATAATGAGATA1302	
RESULT 2			
LOCUS	AR428809	1526 bp	DNA linear PAT 18-DEC-2003
DEFINITION	Sequence 23 from patent US 6642041.		
ACCESSION	AR428809		
VERSION	AR428809.1	GI:40188595	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1526)		
AUTHORS	Chen,J., Feder,J.N., Nelson,T.C., Kyrstek,S.R. and Duclos,F.		
TITLE	Polynucleotides encoding a novel metalloprotease, MP-1		

JOURNAL	Patent: US 6642041-A 23 04-NOV-2003;
FEATURES	Location/Qualifiers
Source	1..1526 /organism="unknown" /mol_type="genomic DNA"
ORIGIN	
Alignment Scores:	
Pred. No.:	1..5e-126
Score:	1385.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	6
US-10-649-273-2_COPY_148_414 (1-267) x AR428809 (1-1526)	
QY	1 MetGluAaIaIaIaLeuThrIleArgLeuThrAspLeuValGluPheProPheLeuVal
Db	1 ATGAGAGGCTATCATCTTACTATATAGGTGACCAAAAGAGTAAATTCCTTTTAAATT
QY	21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu
Db	61 CTTTGAATTTCTGAGGTCACTCTCTGTTGGCATTAAGTTCAAGAGATTTCAGATTTCG
QY	41 LeuLeuGlyIysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg
Db	121 CTTCTTGGAAAGCTTTTGGACATAGACACAGAGTGACATCTTGGACAGAGTGCAAGAGA
QY	61 LeuSerLeuIleIysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlnHis
Db	181 CTTTCTTTAAATTAACCATCCAGAGTCTCTCACCATAGAGTGGGAAAGCCATTAGACAT
QY	81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys
Db	241 TTGGCCAAACAGAGAAATAGATTCTTTTGGACATCAAACTCTTGCATCATCTTAA
QY	101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys
Db	301 AATTGATATTTTCTTTTACTGGACCTTCAACAGCTTACGTATTAATATATGAAAAG
QY	121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla
Db	361 GAAAAAAGAGAGGTATTGAGAGGGGCAAAATCTGTCTTCAGCAGCAGACATTCCTGCC
QY	141 ThrValGlnHisIsthreAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe
Db	421 ACACTACACACACAAATGGCATTCATCTTGGAAAGAACACATCGGGCTTATTCGTTT
QY	161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAl
Db	481 TGTAGCAGAGAGACTGTTTACTCTCAAAATTAATGCGATCTGCTGCATCTGCTGCTC
QY	181 AlaSerAsnPheIleIleArgAlaAlaGluIleLeuThrAsnAlaThrGlnCysThr
Db	541 GCAAGTAACCTTCAATATCCGACAGAGCTTCGAAATTTTAAACAAAGCAACACAGGCACT
QY	201 LeuLeuCysProProProAlaArgLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly
Db	601 TTGTTGTGTCTCTCCACAGCATATGCATGTATATGATTCATTAATTCATGAAATGCT
QY	221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu
Db	661 ATTGAAAGACTAGCTGTGCTGGCATTTTAAACATGACATAGAAAGGATCCGCTATGAA
QY	241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal
Db	721 CCAAAATGTCCTCTTGGAGTAGACATATCAAAAGAAAGTGGAAGCTTCCATAAAAGTA
QY	261 ProGlnLeuLysMetGluIle 267
Db	781 CCACAAATTAATAAATGAGATA 801

RESULT 3
BC011904
LOCUS
DEFINITION Homo sapiens O-6-acyloglycoprotein endopeptidase-like 1, mRNA (cDNA
clone MGC:20293 IMAGE:4121450), complete cds.
ACCESSION BC011904
VERSION BC011904.2
KEYWORDS GI:40225818
SOURCE MGC.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1908)
REFERENCE
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheefer, C.F., Blac, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ueding, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Vitaliano, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Boutard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Stalke, U., Small, D.E.,
Scherer, A., Schein, J.B., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1908)
Strausberg, R.
Direct Submission
Submitted (30-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NHI-MGC Project URL: <http://mgc.nci.nih.gov>
On Dec 19, 2003 this sequence version replaced gi:15080281.
REMARK
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMI)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgcembgrl.nih.gov
Ahter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R.,
Maduro, O.L., Masello, C., Masker, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tougeon, C., Vogt, J.L., Walker, M.A., Wechtery, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, B.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILMI at: <http://image.llnl.gov>
Series: IRAL Plate: 28 Row: 1 Column: 22.
Location/Qualifiers
1. 1908
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:20293 IMAGE:4121450"
/tissue_type="Muscle, rhadomyosarcoma"

US-10-649-273-2_COPY_148_414 (1-267) x BC011904 (1-1908)
ORIGIN
misc_feature
1 MetGUAAlAHIAleuThriLeaTgLeuThraPheValAlGluPheProPheLeuVal 20
545 ATGGAGGCGCTCAAGCACTTCTATTAGTGGACCAATGAAGTAAATTCCTTTTACGTT 604
21 LeuLeuAlIeSerGlyGlyHisCysLeuLeuAlaLeuValAlGlyGlyValSerAspPheLeu 40
605 CTTTGATTTCTGAGAGTCACTGCTGTTGGCAATTAGTTCAGAGATTCAGATTTCG 664
41 LeuLeuGlyLySerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
665 CTTTGGAAAGCTCTTTGGACATAGACACAGGTGACATGCTTGACAGAGTGGCAAGA 724
61 LeuSerLeuIleLyHisPProGlyCysSerThrMetSerGlyGlyLyAlaIleGluHis 80
725 CTTTCTTTAATAAATCATCCAGAGTCTCCACATCATAGTGGGAAGCCATAGAACAT 784
81 LeuAlaLyGlnGlyAsnArgPheHisPheAspIleIleValProProIleuHisAlaLy 100
785 TTGGCCAAACGAAAGATGATTGATTTGACATCAAACTCCCTTGATCATGCTAA 844
101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLyIleIleMetLyLy 120
845 AATTGTGATTTCTTTTACTGACTCAACGCTTACTGATTAATAATAATGAAAAG 904
121 GluLyGlnGluGlyIleGluLyGlnIleLeuSerSerAlaAlaAspIleAla 140
905 GAAAAGAGAGAGATGATGAGAGGGCAATCTGTTCTTCAGACAGACATGTCGCC 964
141 ThrValGlnHisThrMetAlaCysHisLeuValIleArgThrHisArgAlaIleLeuPhe 160
965 ACAGTACAGACACATGCGATGATCTTCTGAAAAGAACATCGGGCTATTCTGTT 1024
161 CysLyGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyVal 180
1025 TGTAAAGCAGAGACTTGTACTCTCAAAATTAAGCATGCTGTCATCTGTCGTC 1084
/clone_1ib="NHI_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. 1908
/gene="OSGEP1"
/db_xref="LocustID:64172"
104. 1348
/gene="OSGEP1"
/codon_start=1
/product="OSGEP1 protein"
/protein_id="AAH1904.1"
/db_xref="GI:15080282"
/db_xref="LocustID:64172"
/translation="MLILRTKGVFPKPKRYEPLRSPNPHPTGLPLKIVLGIET
SCDDPRAAYVDERTGNVLGEAHISQTPVHAKTGIVPPAAQQRHNTQIVQALAS
GVSFSLSAIVATTIKGLALSLGVGSFLQVLGQKKPPIPIHHEALATLRLTK
VAFPEPLVLLISGHCLALVQVSDPLLGKSLIDIPGMVDVARSLIKPECS
MISGKAIETHIAQGNFHDIPKPLHAKNCDPSFYLQPTVDTKIMEKEKEGIEG
GILSSAADIAATVOTHTMAHLVRRTRAILFCQRDLPPNNNAVLVAGSVASNPYIR
RLLEITNNATCTLICPPRLCTDNGIMTANCIERLRAGLGIHIDIEGIRYKCPPL
GVDISKEVGRASIKYPLKMEY"
215. 1111
/gene="OSGEP1"
/note="Peptidase M22; Region: Glycoprotease family"
/db_xref="CD:pfam00814"

QY 181 Alaseaenpheyrlleaargalaleuuglileuethrasnalathrglncysthr 200
DB 1085 GCAAGTAACCTCTATATCCGAGAGCTCTGGAAATTTTAAACAAGCAACAGTGCAT 1144
QY 201 LeuleucyspProProProargleucysThraspaanglyllemetilleatrpangly 220
DB 1145 TTGTTGTGTCCTCCCTCCAGACTAGCACTGATATATGCAATTATGCAAGATGAT 1204
QY 221 lilegluargleuargalaglyleuglylileuuhiaspillegluglylleaarglyrglu 240
DB 1205 ATTGAAGACATACGCTGCTGGCATTTTACATGACATGAGAGGATCCGCTATGAA 1264
QY 241 ProlyscysProleuglyvalaspilleserlysgluvalglglualaserilleysval 260
DB 1265 CCATAATGCTCTCTTGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCCATTAAGTA 1324
QY 261 Progluleuylsmetgluile 267
DB 1325 CCACAATTAATAATGAGATA 1345
RESULT 4
AR428803 AR428803 2197 bp DNA linear PAT 18-DEC-2003
LOCUS Sequence 1 from patent US 6642041.
DEFINITION AR428803
ACCESSION AR428803
VERSION AR428803.1 GI:40188589
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2197)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, Mp-1
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;
FEATURES
source 1..2197
location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,33e-126 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AR428803 (1-2197)
QY 1 MetglualahialaleuThrlleaargleuThrasnlysvaIglupheProphleuVal 20
DB 672 ATGAGGCTCATGCACTTACTATTAGCTTGACCAATTAAGTAATTCCTTTTACTT 731
QY 21 leuleuilesarglygylhiEscyleuenualeuvalGlnglyvalaserapheleu 40
DB 732 CTTTGTATTTCTGGAGGTCACTGTCTGTGGCATTTGATCAAGAGTTTCAATTTTCTG 791
QY 41 leuleuglylysserleuaspillealaprogllyaspmetleuaspilyvalaIaargarg 60
DB 792 CTTCTTGAAAGTCTTTGGACATACCAAGGTGACATGCTTGACAAAGTGGCAAGADA 851
QY 61 leuSerleuilelyehiaproglucysserThrmetSerglylyvalaIlegluhis 80
DB 852 CTTTCTTAATAAATCAATCCAGAGTCTCCACATGAGTGTGGGAAAGCCATAGAACAT 911
QY 81 leuulalyvgInglyasnarghphelaspilleysproProleuhishialyvs 100
DB 912 TTGGCCAAACAGAAATGATTTGATTTGACATCAACCTCCCTTGATATGCTTAA 971
QY 101 AsnCyaspPheSerPheThrglyleuglnhisvalIThrAsplysIlelleuetylyvs 120
DB 972 AATTGTGATTTTCTTTTACTGACTTCAACGTTACTGATTAATAATATGAAAAAG 1031

QY 121 GluylsGluGluGlylileglubysglynileuSerSerAlaIaaspilleaIa 140
DB 1032 GAAAAAGAGAAAGATTTAGAAAGGGCAAAATCTCTCTTACAGCAACATGCTGCC 1091
QY 141 ThrValGlnhisThrMetAlaCysnileuVallyvargThrhiasrgalailleupe 160
DB 1092 ACAGTACAGCACACATAGGCATGTCATCTTGAAAAAGAAACATCGGGCTATTCTGTT 1151
QY 161 CyslysglnargaspileuLeuProGlnaenanaIaIaIeuvAlaIaserlygylval 180
DB 1152 TGTAAAGCAGAGACTGTTTACTCAAAATATGCAAGTCTGTTGCATCTGCTGTC 1211
QY 181 Alaseaenpheyrlleaargalaleuuglileuethrasnalathrglncysthr 200
DB 1212 GCAAGTAACCTCTATATCCGAGAGCTCTGGAATTTTAAACAAGCAACAGTGCAT 1271
QY 201 LeuleucyspProProProargleucysThraspaanglyllemetilleatrpangly 220
DB 1272 TTGTTGTGTCCTCCCTCCAGACTATGCACTGATATGAGCATTTGATGATGGAATGAT 1331
QY 221 lilegluargleuargalaglyleuglylileuuhiaspillegluglylleaarglyrglu 240
DB 1332 ATTGAAGACATACGCTGCTGGCATTTTACATGACATGAGAGGATCCGCTATGAA 1391
QY 241 ProlyscysProleuglyvalaspilleserlysgluvalglglualaserilleysval 260
DB 1392 CCATAATGCTCTCTTGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCCATTAAGTA 1451
QY 261 Progluleuylsmetgluile 267
DB 1452 CCACAATTAATAATGAGATA 1472
RESULT 5
AR428808 AR428808 1387 bp DNA linear PAT 18-DEC-2003
LOCUS Sequence 21 from patent US 6642041.
DEFINITION AR428808
ACCESSION AR428808
VERSION AR428808.1 GI:40188594
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1387)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, Mp-1
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;
FEATURES
source 1..1387
location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 5,43e-124 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x AR428808 (1-1387)
QY 1 MetglualahialaleuThrlleaargleuThrasnlysvaIglupheProphleuVal 20
DB 465 ATGAGGCTCATGCACTTACTATTAGCTTGACCAATTAAGTAATTCCTTTTACTT 524
QY 21 leuleuilesarglygylhiEscyleuenualeuvalGlnglyvalaserapheleu 40
DB 525 CTTTGTATTTCTGGAGGTCACTGTCTGTGGCATTTGTTCAAGAGTTCAGATTTTCTG 584
QY 41 leuleuglylysserleuaspillealaprogllyaspmetleuaspilyvalaIaargarg 60

D5	585	CTTCTGGAAAGTCTTTGGACATACGACACAGGTGACATGCTTGACAGGTGGCAAGA	644				
Oy	61	LeuSerLeuIleIyehSHeProGluCyserThrMetSerGlyIyValaIleGluHis	80				
D5	645	CTTCTTTAAATAAACATCCAGAGGTCTCCACATGATGCTGGAAAGCCATAGACAT	704				
Oy	81	LeuAlaIySGInGlyAsnArgPheHisIshAspIleIySerProProLeuHisHisAlaIyS	100				
D5	705	TTGGCGAAACAAAGAAATAGATTTTCATTTTGACATCAAACTCCCTTGATCATGCTAA	764				
Oy	101	AsnCyAspPheSerPheTherGlyLeuGlnHisValIthAspIyIleIleMetLysIyS	120				
D5	765	AATGTGATTTTCTTTTACTGGACTTCAACGCTTACTGATTAATTAATGAAGAG	824				
Oy	121	GluIySGInGluGlyIle-----	126				
D5	825	GAATAAGAGAGAGATATTTCTAATATGTAAGATTGAACAGATTAATATTCCTGATTG	884				
Oy	127	-----GluIySGInGlnIleLeuSerSerIa	135				
D5	885	TGCCTAAATAATAGCTGCTCATTTCTGACGATAGAGAAAGGGCAATCCTGTCTTCAGCA	944				
Oy	136	AlaAspIleAlaIaIthValGlnHisIthMetAlaCySHisIleValIyAspThrHis	155				
D5	945	GCAGACATGCTGCGACAGTACAGACACATGATGATCATCTTGTGAAAGAACACAT	1004				
Oy	156	ArgAlaIleIeIuePheCyAsySGInArgAspLeuProGlnAsnAsnAlaValIleuVal	175				
D5	1005	CGGGTATATCTGTTTGTATAGACAGAGACATCTTGTAACCTCAAAATATATGCACTGGTT	1064				
Oy	176	AlaSerGlyIyValAlaSerAsnPheIyIleArgAlaAlaLeuGluIleLeuThrAsn	195				
D5	1065	GCATCTGTGTGTGTCGCAAGTAACTCTTATATCCGACAGCTCTGGAATTTTAACAAC	1124				
Oy	196	AlaIthArgInCySThrIleuCyAspProProAlaGlyLeuCySThrAspAsnGlyIleIleMet	215				
D5	1125	GCAACACAGTGCATCTTGTGTGTCCTCTCCACAGCATATGACATGATTAATGGCATTAATG	1184				
Oy	216	IleAlaIthAsnGlyIleGluArgLeuArgAlaGlyIleLeuHisAspIleGlu	235				
D5	1185	ATTGCATGGAATGGATATGTAAGAAAGCTAGCTGTGGCATTTTACATGACATAGAA	1244				
Oy	236	GlyIleArgIyrgIuProIyCyAspProLeuGlyValaAspIleSerLysGlyIyValGlyIu	255				
D5	1245	GGCATTCGCCCTATGAAACCAAAATGTCTCTTGAGATGACATATCAAAAGAGTTGAGAA	1304				
Oy	256	AlaSerIleIyValProGlnLeuIyMetGluIle	267				
D5	1305	GCTTCATTAATAAGTACCACTAATTAATAATGAGATA	1340				
RESULT 6	HSAA295148	1387 bp	mRNA	linear	PRI 30-OCT-2000		
LOCUS	HSAA295148	1387 bp	mRNA	linear	PRI 30-OCT-2000		
DEFINITION	Homo sapiens mRNA for putative sialoglycoprotease type 2.						
VERSION	AJ295148.1	GI:11071726					
KEYWORDS	metallopeptidase; sialoglycoprotease.						
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						
REFERENCE	Chen J.M., Fortunato M. and Barrett A.J.						
AUTHORS	Mumukshu; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;						
TITLE	Eumalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
JOURNAL	1						
REFERENCE	Cloning and sequencing of a second human putative						
AUTHORS	sialoglycoprotease homologue						
TITLE	Unpublished						
JOURNAL	2 (bases 1 to 1387)						
REFERENCE	Chen J.M.						
AUTHORS	Direct Submission						
TITLE	Submitted (27-OCT-2000)						
JOURNAL	Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,						
FEATURES	UNITED KINGDOM						
FEATURES	Location/Qualifiers						

```

source
1. 1387
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /tissue_type="uterus"
   24..1343
   /function="metalloproteinase of family M22"
   /codon_start=1
   /product="putative sialoglycoprotease type 2"
   /protein_id="CAC14666.1"
   /db_xref="GI:11071727"
   /db_xref="Gene:09H4B0"
   /db_xref="UniProt/TREMBL:Q9H4B0"
   /translation="MLIKTQAGVFFPFSKRVKVPFELRSFNPHBETLFLKIVIGIET
SCDTPAAAVDETNVLEGEALHSQTEVHLKTKGVLPVPAQQLHRENTQRIYQVALSAS
GVSFSDLSAIAITTKIPGLALVGVSGDPLFLGSLQALPDMDKVARLSLTKHPCGST
VERPFLVLLISGGHCLALVQGVSDPLFLGSLQALPDMDKVARLSLTKHPCGST
SKGKALTEHLAKOGNRFHFDIKPRLHKANKDPSFTGLQHTDKIKKKEKGIPLLI
SKVQKQINPGLCLKIAAHFPCRYEKGQILSSADPAATQTHMAHLVYKTRAILFPC
QRDLTPQNNAVLVASGVASNFYIRLAEILTNATQCTLLCPPLRLCTDNGIMTANNG
IERLRGGIGILHDIIEGIRYEKCPGLAVDISKEVEGASIKVQLNME1"

ORIGIN

Alignment Scores:
Pred. No.:      5,43e-124      Length:      1387
Score:          1358.50        Matches:      266
Percent Similarity: 91.10%      Conservative: 0
Best Local Similarity: 91.10%      Mismatches: 1
Query Match:     98.09%        Indels:       25
DB:              9            Gaps:         1

US-10-649-273-2_COPY_148_414 (1-267) x HSA295148 (1-1387)

Oy      1 MetGluAlaHisAlaLeuThrTilAArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      465 ATGAGAGGCTCAGCACTTACTATTAGGTGACCAATAAGTAGAATTTCTTTTACTT 524

Oy      21 LeuLeuLieserGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      525 CTTTGGATTCTGGAGGCTCAGCTGCTGTGGCATTTAGTTCAAGAGGATTTCAGATTTCTG 584

Oy      41 LeuLeuGlyLysSerLeuAspLilAlaLProGlyAspMetLeuAspLysValAlaArgArg 60
Db      585 CTTCTTGGAAAGCTCTTTGGACATGACACCAAGTGTACATGCTTGGACAGAGTGGCAAGAGA 644

Oy      61 LeuSerLeuLilLysHisAspGluLysCysSerThrMetSerGlyGlyValAlaLleGlnHis 80
Db      645 CTTTCTTTAATAAACAATCCAGAGGTGCTCCACCAAGAGTGGGAAAGCCATAGAGCAT 704

Oy      81 LeuAlaLysGlnGlyAsnArgPheHisPheAspLilLysProProLeuHisHisAlaLys 100
Db      705 TTGGCCAAACAAGAAATAGATTTCATTTTACATCAACAACTCCCTTGCATCATGCTTAA 764

Oy      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysLilLleMetLysLys 120
Db      765 AATGTGATTTTCTTTTACTGTGACCTTCAACAGCTTCTGATTAATAATATGAAAGAAG 824

Oy      121 GluLysGluGluGlyLilLe----- 126
Db      825 GAAAAAAGGAGGAAGTATATTTCTAATTAGTAAAGTTCAACAGATTAATATTTCTGGATTG 884

Oy      127 -----GluLysGlyGlnLilLeuSerSerAla 135
Db      885 TGCCATAAATAAGCTGCTCATTTTCTGCAGGATATGAGAAAGGGCAATCTCTTTCAGCA 944

Oy      136 AlaAspLilLeuAlaLThrValGlnHisThrMetAlaCysHisLysValLysArgThrHis 155
Db      945 GCACACATGTCGTCGACAGTACACACACATGCGCATGTCATCTTGTGAAAGAACAACAT 1004

Oy      156 ArgAlaLilLeuLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
Db      1005 CGGGCTATTTCTGTTTGTAAAGCAGAGACATTTGTACTCTCAAAATTAATGACGATCTGGTT 1064

```


Oy	176	11a5eerglygVla1A1a5eAspNheTyr11e4rgarGla1a5Glu11e4euThrAsn	195		
Db	1065	GCATCTGATGATGTCGCAAGTAACCTTATATCCGACAGCTCTGGAAATTTTAAACAAAC	1124		
Oy	196	AlaThrGlnCystrhLeuLeuCyseProProProAArgLeuCystrhAspAsnGly11e4et	215		
Db	1125	GCAACACAGTGCACTTTGTGTGTCTCTCCAGACATGATGACATGATTAATGGCATTTATG	1184		
Oy	216	11e4a1aTTPaasnGly11e4GluAArgLeuAArgLaGly11e4Gly11e4H1aAsp11e4Glu	235		
Db	1185	ATTTCATGATGAATGGATTAAGAAAGACTAGTGTGGCTGGCAATTTTAAACATGACATGAGA	1244		
Oy	236	Gly11e4ArgTyrGlnProLysCyseProLeuGlyValaAsp11e4SerLysGluValaGly11e4	255		
Db	1245	GGCATCCGCTATGAACCAAAATGTCCTTTGAGATGACATATCAAAAGAAAGTTGGAGAA	1304		
Oy	256	AlaSer11e4LysValaProGlnLeuLysMetGlu11e4	267		
Db	1305	GCTTCCATTAAGATACCAATTTAAATGAGATTA	1340		
RESULT 7					
AX664697	AX664697	1245 bp	DNA	linear	PAT 22-MAR-2003
LOCUS	Sequence 6 from Patent WO02074960.				
DEFINITION	AX664697				
ACCESSION	AX664697.1	GI:29164457			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Carartinhia; Homnidae; Homnide; Homo.				
REFERENCE					
AUTHORS	1				
TITLE	Leibdy,K.R., Kapeller-libermann,R. and Glucksmann,M.				
JOURNAL	38550, 28472, 5495, 65507, 81588 and 14354 methods and compositions				
	of human proteins and uses thereof				
	Patent: WO 02074960-A 6 26-SEP-2002;				
	Millennium Pharmaceutical, Inc. (US)				
FEATURES					
source	Location/Qualifiers				
	1..1245				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
ORIGIN					
Alignment Scores:					
Pred. NO.:	5.14e-123	Length:	1245		
Score:	1348.00	Matches:	260		
Percent Similarity:	98.50%	Conservative:	3		
Best Local Similarity:	97.38%	Mismatches:	4		
Query Match:	97.33%	Indels:	0		
	6	Gaps:	0		
US-10-649-273-2_COPY_148_414 (1-267) x AX664697 (1-1245)					
Oy	1	MetGluAlaH1a5a1aLeuThr11e4ArgLeuThrAsnLysValaGluPheProPheLeuVal	20		
Db	442	ATGAGGCTCATGACATTAATTAAGTTGACCAATTAAGTAAGTAATTTCCCTTTTATGTT	501		
Oy	21	LeuLeu11e4SerGlyGly11e4CysLeuLeuAlaLeuValaGlnGly11e4ValaSerAspPheLeu	40		
Db	502	CTTTTGATTTCTGGAGGTCACATGTCGTGTGGCATTTAGTTCAAGAGATTTCAGATTTCTGTG	561		
Oy	41	LeuLeuGlyLysSerLeuAsp11e4AlaProGlyAspMetLeuAspLysVala1aArgTyr	60		
Db	562	CTTCTTGGAAGTCTTTTGACATGACACAGTGAACATGCTTGAACAAGTGGCAAGAGA	621		
Oy	61	LeuSerLeu11e4LysH1a5ProGluCysSerThrMetSerGlyGlyLysVala11e4GluH1a5	80		
Db	622	CTTTCTTTAATAAAACATCCAGAGTGTCCACATGACATGAGTGGGAAAGCATAGACAT	681		
Oy	81	LeuAlaLysGlnGlyAsnAArgPheH1a5PheAsp11e4LysProProLeuH1a5AlaLys	100		
Db	682	TTGGCCAAACAGGAAGAAATGATTTTCATTTTAAACATCAACCTCCCTTGATCATATGCTTAA	741		

Oy	101	AamCysaapbpeSerPethnGlyLeuGlnHisValThrAspLysIleIleMetLysVal	120
Db	742	AAATGTATATTTCTTTTACTGACCTTCAACACCTTACTATATAAATAATGAAACAGG	801
Oy	121	GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIleAla	140
Db	802	AAAAACAGAGGAGGTATTGAGAAAGGGCAATCTGTCCTTACAGCAGCAGCATTTGCTGCC	861
Oy	141	ThrValGlnHisSerThrMetAlaCysHisLeuValIleAspArgThrHisAlaGlnAlaIleLeuPhe	160
Db	862	ACACTACAGCAGCAACAATGAGCATGTCATCTTGTAAGAAAGAACACATCGGGCTATTCTGTTT	921
Oy	161	CysLysGlnArgAspLeuLeuProGlnAspAsnAlaValLeuValAlaSerGlyGlyVal	180
Db	922	TGTATGACAGAGAGCTTGTTACTTCAATAATATACAGTACTGCTGTGATCTGGTGGTGTCTC	981
Oy	181	AlaSerAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr	200
Db	982	GCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTAAACAACGCAACAGTGCACT	1041
Oy	201	LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleTyrAsnGly	220
Db	1042	TTGTTGTGCTCTCTCCCAAGCTATGACATGATATATGCACTTAAGATTGACATGAAATGCT	1101
Oy	221	IleGluArgLeuArgAlaGlyLeuGlyIleIleuHisAspIleGluGlyIleArgTyrGlu	240
Db	1102	ATTGAAAGACTACGCTGCTGGCTTGCGCATTTTATCATACATAGAGGCAATCCGTTATGAA	1161
Oy	241	ProLysCysProLeuGlyValAlaAspIleSerLysGluValGlyGlnAlaSerIleLysVal	260
Db	1162	CCAAATATGTCCTCTTGAGATAGACATATCAAAAGAAATTGAGACAGCTTCATTAAGTA	1221
Oy	261	ProGlnLeuLysMetGluIle	267
Db	1222	CCACAATTAAAAATGAGAGATA	1242
RESULT 8			
AX664695	LOCUS	AX664695	1820 bp DNA linear PAT 22-MAR-2003
DEFINITION	Sequence 4 from Patent WO02074960.		
ACCESSION	AX664695		
VERSION	AX664695.1	GI:29164455	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	BuayaVote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1	Leibyl,K.R., Kapeller-liberman,R. and Glucksmann,M.	
AUTHORS	36550, 28472, 5495, 65507, 81588 and 14354 methods and compositions		
TITLE	of human proteins and uses thereof		
JOURNAL	Patent: WO 02074960-A 4 2c-SRP-2002;		
	Millennium Pharmaceuticals, Inc. (US)		
FEATURES	Location/Qualifiers		
source	1..1820		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	146..1390		
	/note="unamed protein product"		
CDS	/codon_start=1		
	/protein_id="CAD80044.1"		
	/db_xref="GI:29164455"		
	/translation="MLILTRTACVPRPKSRKYVEPLRSPFNPHGTLFLKIVLGIET		
	SCDDTAAVVDEGTAVLGEAHSCTEVLHKTGGIVPPAQQLRENIQRIYQELASMS		
	GVSPSDLSAIAATTIKPGLASLISLGVGSFSIQVLQVKPKPTPIIHNEAHVLTIRLNK		
	VEPFVLTLISGGICLALVQGVSDPLILKSLSDIAPGMDLKYARLSIKIPBCST		
	MSGGALIEHLAKQGNRPHDIPKPLIHAKCDFSPGILQCHTDKNENRNRQOEGIEKG		
	QISSADIAATVQHTWACHLVKRTIRAILIFCORDDLPNNKLVLASGVASNPYRL		
	RALEILTNATQCITLCPRPRLCTDNGIMIMAMNGIERLRAGLGIHLIDIBGIRYEPKPL		
	GVDLSKVEGASIKVPOLKMET"		
ORIGIN			

Alignment Scores:

Pred. No.: 8,17e-123 Length: 1820
 Score: 1348.00 Matches: 260
 Percent Similarity: 98.50% Conservative: 3
 Best Local Similarity: 97.38% Mismatches: 4
 Query Match: 97.33% Indels: 0
 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x AX664695 (1-1820)

QY 1 MCGUUAHAIHAIAlaLeuThi1leArgLeuThraNlySvalGluPheProPheLeuVal 20
 DB 587 ATGAGGCTCAGTACACTACTATTAGTTGACCAATTAAGTAGAATTTCTTTTACTG 646
 QY 21 LeuLeu1leSerg1yG1yHicSyeLeuLeu1aLeuValG1yG1yValSerAspPheLeu 40
 DB 647 CTTTGTATTTCTGAGGCTCAGTCTGTGTGCTATTAGTTCAAGAGGTTTCAGATTTCTG 706
 QY 41 LeuLeuG1ySerg1ySerg1yLeuAsp1leAlaProG1yAspMetLeuAsp1yValAlaArg 60
 DB 707 CTTCTTGAAAGTCTTTTGACATACACAGTGTGACATCTTGACAGAGTGGCAAGAGA 766
 QY 61 LeuSerLeu1lelyS1sProG1yCysSerThraMetSerg1yG1yLySAla1leG1uH1s 80
 DB 767 CTTTCTTAATAAACAATCCAGAGTGTCCACATGAGTGTGGGAAAGCCATTAGACAT 826
 QY 81 LeuAla1ySG1yG1yAenArgPheH1sPheAsp1lelySProProLeuH1sAla1yS 100
 DB 827 TTGCTCAACAAGAGGATATTGAGAGGGAATTTGACATCAACCTCCCTTCATCATCTTAA 886
 QY 101 AsnCysAspPheSergPheThG1yLeuG1nH1sValThraP1yS1le1eMetLyS1yS 120
 DB 887 AATTGTATTTTCTTTTCTGACCTTCAACGTTACTGATTAATAAATAAGAAACAGG 946
 QY 121 Glu1ySG1yG1y1leG1uLyS1yG1n1leLeuSerg1yAla1leAla1y 140
 DB 947 AAACAAGAGGAGGATATTGAGAGGGAATTTCTGTTGACGACGACGACATTTCTGCC 1006
 QY 141 Th1ValG1nH1sThraMetAlaCysH1sLeuVal1ySArgThra1leArgAla1leLeuPhe 160
 DB 1007 ACAGTACAGACACAAATGCAATGTCATCTTGTGAAGAAACACATCGGCTATTCTGTTT 1066
 QY 161 Cys1ySG1yArgAsp1leuProG1nAspAsn1aVal1leuValAla1leSerg1yG1yVal 180
 DB 1067 TGTAAAGAGAGAGCTGTGTAACCAAAATATGACAGTGTGCTGCTGCTGCTC 1126
 QY 181 AlAserAspPheTyrl1eArgArgAlaLeuG1u1leLeuThraAsn1aThraG1nCyThr 200
 DB 1127 GCAAGTAACTTCTATATCCGAGAGCTGTGAATTTTAAACAAAGCAACAGTGCAC 1186
 QY 201 LeuLeuCySProProArgLeuCyThraSpsang1y1leMet1leAla1leThraSng1y 220
 DB 1187 TTGTGTGTCTCTCTCCAGACTATGACCTATATGCACTTATGATTTGATGAGATGCT 1246
 QY 221 1leG1uArgLeuArgAlaG1yLeuG1y1leLeuH1sAsp1leG1uG1y1leArgT1yG1u 240
 DB 1247 ATGAAAGACTTACCTGTGCTGTGGCACTTTTACATGACATAGAGGATCCGCTATGAA 1306
 QY 241 ProLySAspProLeuG1yValAsp1leSerg1yG1uValG1yG1uAla1leSerg1yVal 260
 DB 1307 CCAAAAGTCTCTCTGAGTAGACATATCAAAAGAGTTGAGAGGCTTCAATAAGTA 1366
 QY 261 ProG1nLeuLySmetG1u1le 267
 DB 1367 CCACAATTAATAATGAGATTA 1387

RESULT 9
 BC058172 EC059172 1644 bp . mKNA linear ROD 08-OCT-2003
 LOCUS Mus musculus cDNA clone MGC:67870 IMAGE:5012054
 DEFINITION BC058172
 ACCESSION BC058172
 VERSION BC058172.1 GI:34849663

KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS

MGC.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1844)
 Strausberg,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Klausner,R.D., Collins,F.S., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahney,J., Helton,B., Kettner,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzywinski,M.I., Skalek,U., Smallus,D.B.,
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 MEDLINE
 PUBMED
 12477932
 2 (bases 1 to 1844)
 Strausberg,R.
 Direct Submission
 Submitted (15-SEP-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgasb@remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Contact: nisc.mgc@nih.gov
 Web site: <http://www.nisc.nih.gov/>
 Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
 Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Latic,P., Legaspi,R.,
 Maduro,Q.L., Mastello,C., Maskerl,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Stancirpop,S., Thomas,P.J., Touchman,J.W.,
 Tsougenou,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,E.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
 Series: IRAK Plate: 123 Row: d Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 21512463.
 Location/Qualifiers
 1. 1844
 /organism="Mus musculus"
 /mol_type="cDNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="MGC:67870 IMAGE:5012054"
 /tissue="type="mammary tumor metastasized to lung. Tumor
 arose spontaneously from a senescent normal mammary
 (clonal) outgrowth infected with the virus MMTV."
 /clone_id="NCT CGAP_Lu29"
 /lab_host="DH10B"

FEATURES source

CDS

/note="Vector: pCMV-SPORT6"
222..1466
/codon_start=1
/product="unknown (protein for MGC:67870)"
/protein_id="AAH58172.1"
/db_xref="GI:34849664"
/translation="MMLRRTAGAI PKPKSKVYGLRFSVHPRTLSCHKLVIGIET
SCDPTGAADVETGNVGBALHSGOTVLTGTGIVPAPAOQRENIORIVETLSAS
RIPPSDLSATATTKPGILALSLGVGLSPSLQVLRKKPPIPIHMEHALITRTNK
VEPPVLVLSGGHCLALVQVSDPFLKGLSLDIPQMDLKVARRSLIKHPECT
MSGKALBHLAKDNRRFHFTINPPQNKNDPFTGLQHTDKLITKKEEGELKQ
QILSSADIAAVQHAACHLAKRTKRAILFCQKNLISPAVAIVSGVAGSNIYIR
KALEIVANATQCTCLCEPPRLCTDNGIAMIWNGIERLRAGGLHVDVDIRPEPCPL
GIDISREVAEAAIVPRILKML"

misc_feature
333..1397
/note="QRT7; Region: Metal-dependent proteases with
possible chaperone activity [Posttranslational
modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

ORIGIN

Alignment Scores:
Pred. No.: 2,76e-111 Length: 1844
Score: 1231.00 Matches: 233
Percent Similarity: 92.88% Conservative: 15
Best Local Similarity: 87.27% Mismatches: 19
Query Match: 88.88% Indels: 0
DB: Gaps: 10

US-10-649-273-2_COPY_148_414 (1-267) x BC058172 (1-1844)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluProPheLeuVal 20
DB 663 ATGAGAGCTCAGCAGCTAGTAATTAAGCTCACCAATTAAGTAAGATTTCTTTCTAGTT 722
QY 21 LeuLeuIleSerGlyGlyIleCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 723 CTTTGTATTTTGGCGGCTCCTGCTGTGGCATTAAGTCGCAAGGTGTTTCCATTTCTCTG 782
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB 783 CTCTTGGGAAGCTCTTGGCATAGCGCAGCGCATGCTTGAACAAGTGGCAGAAGA 842
QY 61 LeuSerLeuIleLeuHisProGlnCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 843 CTTTCTTTAAACAACATCCAGAAATGTTCTCAATGAAGTGGGAAAAGCTATAGAAAT 902
QY 81 LeuAlaIleGlnGlyAsnArgPheHisAspIleLeuProPheLeuHisAlaIleVal 100
DB 903 TTGGCCAAAGCGAATATGATTCATTTTACTATCAATCCACTATGACAGATCTAAG 962
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetLeuHis 120
DB 963 AATTGCCATTTTCTTTTACGGGACTTCAACATATTATCGATTAAGCTATATAACACAG 1022
QY 121 GluIleGluGluGlyIleGluIleGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 1023 GAAAAAGAGAGGCAATTCGAGAGGGGCAAAATCCTGTATCAGCTGACGACATTCCTCT 1082
QY 141 ThrValGlnHisThrMetAlaCysHisIleValIleValIleValIleValIleVal 160
DB 1083 GCGGTATCAGCATGCAACAGCGTGCACCTTGCAGAAAAGAACACATCGTGTATGTTT 1142
QY 161 CysIleGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyVal 180
DB 1143 TGCAGAGCAAAAAATTTGCTATTCACGCTAAGCATTAATTAAGTGTATTCGAGAGTGT 1202
QY 181 AlAserAsnPhetIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
DB 1203 GCAAGTAATCTGTATCCGAAAGCATTCGAAATTCGCGAAATGACAGACGAGTGCACG 1262
QY 201 LeuLeuCysProProPheArgLeuCysThrAspAsnGlyIleMetIleAlaIleThrAsn 220

DB 1263 TTGCTGTGTCCCTCTCAAGACTGTGCACTGCAAAATGCATCATGATTCAGTAATGGA 1322
QY 221 ILeGluArgLeuArgAlaGlyLeuGluIleLeuHisAspIleGlnGlyIleArgTyrGlu 240
DB 1323 ATTTGAAGATTAAGCTGTGGCTGTGGCTGTTTAACTAGATGATGAGAACATCCGATATGA 1382
QY 241 PolyS(CysProLeuGlyValAlaIleSerIleSerGlyValAlaGlyValAlaSerIleVal 260
DB 1383 CCAAAATTCCTCTTGGATATACATATTCACAGAGAGTTGCGAAGCTCCATTAAGA 1442
QY 261 ProGlnLeuIleMetCysIle 267
DB 1443 CCGCATTAATAATGACACTT 1463

RESULT 10
BC038910
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1017)
Strasberg,R.L., Fellingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Schenker,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Scheffer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaez,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,B.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Keltman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallus,D.E.,
Schnerch,A., Schein,J.B., Jones,S.J., and Watra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 1017)
Strasberg,R.
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bowder, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Matheson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saedi, Jacqueline
Schein, Diane Smilus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,

George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLM at: <http://image.llnl.gov>
Series: IRAX Plate: 86 Row: F Column: 12.
Location/Qualifiers

FEATURES

Source
1..1017
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053559"
/tissue_type="Liver, normal, 5 month old male mouse."
/clone_id="NCI COAP L19"
/lab_host="MDH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.: 6,53e-111 Length: 1017
Score: 1224.00 Matches: 233
Percent Similarity: 92.51% Conservative: 14
Best Local Similarity: 87.27% Mismatches: 20
Query Match: 88.38% Indels: 0
DB: Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x BC038910 (1-1017)

1 MetGluAhiAalaleuThrlaargleuThraenlyValGluPheProPheleuVal 20
26 ATGAGGCTCAGCAGCATATTAAGCTCAACATTAAGATTTCTTTTAAATT 85
21 LeuLeuileSerGlyYhiScyeLeuLeuAaleuValGlnGlyValSerAphleu 40
86 CTTTGAATTCCTGGCGGCTGCTGCTGTCATTAAGTCACAGGTGTTCCGATTTCTG 145
41 LeuLeuGlyVserSerleuAapllleAalProGlyAapMetleuAaplyeValAalarg 60
146 CTCCTTGGAGTCTTTGGACATAGCACAGCAGCATGCTTGAAGGTGGCAAGAGA 205
61 LeuSerleuileYshiSproGluCyserThrmetsSerGlyGlyAalaleuVal 80
206 CTTTCTTAATCAACAATCCAGAAATGTTCTAACAATGAGTGGGAAAAGCTATAGA 265
81 LeuAalalySGlnGlyAasnargPhehisPheAapllleYsProProleuHiehisAly 100
266 TTGGCCAAAGACGGAATAGATTCATTTTACTAATCCACTATCCAGATGCTAAG 325
101 AaNCyAaPheSerPheThrglyLeuGlnhisValThraaplyeValleuMetCyaly 120
326 AATTGCAATTTTCTTTCACGCGACTTCAACATATTACTGATTAAGCTAATAACACA 385
121 GluYsglGluGlyYlleglulysGlyGlnileuSerSerAalAapllleAalale 140
386 GAAAAAAGAGAGCATTTGAGAGAGGGAATCTCTGATCATGCTGCAGCATTCCTG 445
141 ThrValGlnhisThrmelAaCyehisLeuVallySargThrhieAalaleuPhe 160
446 GCGGTACAGCATGCAACAGCGTGCACCTTCGAAAAAAGACATCGGCTATTCGTT 505
161 CyselysGlnAagAapleuPheProGlnAasnAalAalValleuValAalSerGly 180
506 TGCAAGAGAAAAATTTGCTCTCTCCAGCTAACGAGATTAAGTGTATCTGGAGGT 565
181 AlaserAapheYrilearAalaleuGlnileuThraaplyeValleuMetCyaly 200
566 GCAAGTAACTTGTATCATCCGAAGCATTTGAGAAATTTGCGAAATGCAACGAGTGA 625
201 LeuLeuCySproProAargleuCyserThraSasnGlyYllemetleAalAalarg 220
626 TTGTGTGTCACTCCAGATCTGTGCACTGCAATGCAATGCAATGCAATGCAATG 685
221 TleGluAargleuAargAalGlyleuGlyYlleglulysAapllleGluGlyYl 240

|||||
686 ATGAAAGATTAAGTCTGCTGGCTTTTACATGATGTAGAGACATCCGATATGAA 745

241 ProlyCySproleuGlyValAapllleSerlySGlnValGlyGlnAalSerleYsVal 260
746 CCAAAATGCTCTTGGAGTACATATCCAGAGAGTTCAGAGAGTTCAGAGAGTTC 805

261 ProGlnleuYsMetGluile 267
|||||
806 CCGCATTAATAATGCACTT 826

RESULT 11
AX713716
LOCUS AX713716 2208 bp DNA linear PAT 15-APR-2003
ACCESSION Sequence 400 from Patent EP1293569.
VERSION AX713716.1 GI:23888642
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Negahari,K. and
Masuho,Y.
Full-length cDNAs
Patent: EP 1293569-A 400 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
LOCATION/Qualifiers
1..2208
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 1.57e-108 Length: 2208
Score: 1204.00 Matches: 239
Percent Similarity: 89.51% Conservative: 0
Best Local Similarity: 89.51% Mismatches: 4
Query Match: 86.93% Indels: 24
DB: Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x AX713716 (1-2208)
1 MetGluAhiAalaleuThrlaargleuThraenlyValGluPheProPheleuVal 20
785 ATGAGGCTCAGCAGCATATTAAGTGTGACCAATTAAGATTTCTTTTAAATT 844
21 LeuLeuileSerGlyYhiScyeLeuLeuAaleuValGlnGlyValSerAphleu 40
845 CTTTGAATTCCTGGAGTCACTGCTGTTGGCATTAAGTTCAGAGCTTCAAGTTTCTG 904
41 LeuLeuGlyVserSerleuAapllleAalProGlyAapMetleuAaplyeValAalarg 60
905 CTTCTTGAAGTCTTTGGACATAGCACAGGTGACATGCTTGAACAGGTGGCAAGAGA 964
61 LeuSerleuileYshiSproGluCyserThrmetsSerGlyGlyAalaleuVal 80
965 CTTCTTGAATTAACAATCCAGAGTCTCCACCATGAGTGGTGGAAACCATTAAGAT 1024
81 LeuAalalySGlnGlyAasnargPhehisPheAapllleYsProProleuHiehisAly 100
1025 TTGGCCAAAGACGGAATAGATTTCTAATTTGACATCAACCTCCCTTGCATCTGTA 1084
101 AaNCyAaPheSerPheThrglyLeuGlnhisValThraaplyeValleuMetCyaly 120
1085 AATTGATTTTCTTTTATCTGCACTGCAACGTTTACTGATTAATAATTAATGA 1144
121 GluYsglGluGlyYlleglulysGlyGlnileuSerSerAalAalAapllleAalale 140

Db 1145 GAAAAAGGAGGATTTAGAGAGGCGAAATCTGCTTCAGCAGACATGTCGCC 1204
Qy 141 ThrValGlnHisThrMetAlaCysHisIleuValIlyAsrGlyHisArgAlaIleuPhe 160
Db 1205 ACAGTACGACACACATGCGATGTCATCTTGTAAGAAAGAACACATCGGCTATTCGTGT 1264
Qy 161 CysIysGlnArgAspIleuPheProGlnIbnAsnAlaValIleuValAlaSerGlyGlyVal 180
Db 1265 TGTAAAGCAGAGAGACTTGTACTCTCAAAATATATGCAAGTACTGTTGCACTGTGTGTC 1324
Qy 181 AlAserAsnPheTyrIleArgArgAlaIleuGluIleuThrAsnAlaThrGlnCysThr 200
Db 1325 GCAAGTACCTTCTGTATCCGACAGCTCGGAAATTTTAAACAAGCAGACAGTGCACT 1384
Qy 201 LeuIleuCysProProProArgIleuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
Db 1385 TTGTTGTCCTCTCTCCGACACTATGCACTGATTAATGCGATTATATATGCA----- 1435
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
Db 1435 ----- 1435
Qy 241 ProIysCysProLeuGlyValAspIleSerIysGlyValGlyAlaSerIleIysVal 260
Db 1436 ---TGATGTCTCTTGAGATGACATATCAAAAGAGTTGAGAGAGCTTCCTCAATAAGTA 1492
Qy 261 ProGlnIleuIysMetGluIle 267
Db 1493 CCACATTAATAAATGAGAGATA 1513
RESULT 12
AK055441 2208 bp mRNA linear PRI 30-JAN-2004
LOCUS AK055441
DEFINITION Homo sapiens cDNA FLJ30879 fis, clone FBRRA2004592, highly similar
to Homo sapiens mRNA for putative sialoglycoprotease type 2.
ACCESSION AK055441
VERSION AK055441.1 GI:16550166
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Oca,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Odayashi,M., Nishi,T., Shibahara,T., Tanaka,K.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shiratori,A., Sudo,H., Hosori,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
Tanikawa,M., Yamazaki,M., Niomiya,K., Ishibashi,T., Yamashita,H.,
Muraoka,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yosida,M., Hotta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Mueshino,K., Yuuki,H., Oshima,A.,
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
Goto,Y., Shinkai,F., Wakebe,H., Hishiguchi,H., Watanabe,T.,
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
Oshiki,K., Hiraoka,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
Oyama,M., Hara,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
Nagase,T., Nomura,N., Kikuchi,H., Maehuo,Y., Yamashita,R.,
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.

TITLE Complete sequencing and characterization of 21,243 full-length
human cDNAs
JOURNAL Nat. Genet. 36 (1), 40-45 (2004)
PUBMED 14702039
REFERENCE
AUTHORS Nishi,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakami,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Maehuo,Y., Nagai,K. and Isogai,T.
NEBO human cDNA sequencing project
Unpublished
3 (bases 1 to 2208)
REFERENCE
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEBO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.
FEATURES
source
1..2208
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FBRRA2004592"
/tissue_type="brain"
/clone_1b="FBRRA2"
/dev_stage="fetus"
/note="cloning vector: pME186FL3"
344..1438
/note="unamed protein product"
/codon_start=1
/protein_id="BAB70923.1"
/db_xref="GI:16550167"
/translation="MLILKTAGVFPKPSKRYVEFLRSFNFPGTLFLHKIVLGIET
SCDDPRAAVDPETGVNLGSAHSQTEVHLKRGIVYDPAOCLHRENIQIVDPAISAS
GVSPEDLSIAITTPGLALSLGVSLSLQVGVKKPFIPIHHEBALITRLTNK
VSPPLVLIISGHLALVQGVSDFLGLKSLDLPAMDVKVARRPLIKRPEST
MSGGAIEHLAKQGRFHEDIPPLIHAKNCDPFGIOLVTDKIIIMKEKEGIEKG
QILSSAADIATVQRMACHIVKRTFRALIFPCORULLPQNNALVAVSGVANSFCIR
RALELTNATQCTLCCPPRLCTDNGIMIA"
ORIGIN
Alignment Scores:
Pred. No.: 1,57e-108 Length: 2208
Score: 1204.00 Matches: 239
Percent Similarity: 89.51% Conservative: 0
Best Local Similarity: 89.51% Mismatches: 4
Query Match: 86.93% Indels: 24
DB: Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x AK055441 (1-2208)
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIlyValGluPheProPheLeuVal 20
Db 785 ATGAGAGGCGTCAAGCACTTACTATTAGTTGACCAATTAAGTGAATTCCTTTTATGATT 844
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValAspSerPheLeu 40
Db 845 CTTTGATTTCTGAGAGTCACTGCTGTGTGGCAATTAAGTTCAGAGATTTTCG 904
Qy 41 LeuLeuGlyIlySerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArgArg 60
Db 905 CTTCTTGAAATCTTTTGACATAGACACAGGTGACATCTTGACAAAGTGGCAAAAGA 964

Oy		61	euserleuilielyshisProglucysseerThrmetserygilyvalaillegluhis	80
Db		965	CTTCCTTTAAATMAACATCCAGAGTCTCCACCAGTAGGTGGGAAAAGCATTGAACAT	10244
Oy		81	LeuAlalygsingjlyasnarPhenishaphesapiellysProProleuhiishiAlays	100
Db		1025	TTCGCCAACAAGGAAATAGAATTCATTGATTTGACATCAACCTCCCTTGATCATGCATAA	108
Oy		101	AenCySaapPhreserPetherThglyleuenglhisValIThrAsplylIellImethyllys	120
Db		1085	AAATGTGATTTTCCTTTACTCGAACCTTAACAACGTTACTGATMAAATATATATGAAAAAG	11444
Oy		121	GluyesyglugluglylleGlubysgiylintileuSerSeralalAspilleaiala	140
Db		1145	GAAAAGAGCAAGATGATTGAGAGGGCAAATCCCTGCTTCAGCAGCAACATTCCTGCC	120404
Oy		141	ThrValgluhisthmecaleCyshiisleuVallysargthrhiIsargalaileleudhe	160
Db		1205	ACAGTACAGCACACATGATGCATGTCATCTTGAAAAAACAACATCGGGCTATTCTGTTT	12644
Oy		161	CyelysglnaagapleuleuProglunamemalavalleuValAlaserilyglyVal	180
Db		1265	TGTAAAGCAAGAAGACTGTATCCCAAAATATATGATGATCGTTGACATCGTGGTGTG	13244
Oy		181	AlasearAmphenylrilleArgArgAlaleuenglileleuthraAmAlathrgIncyThr	200
Db		1325	GCAAGTAACTTCTGTATCCGACAGACTCTGAAATTTTAAACAAAGCAACAGTGCACCT	13844
Oy		201	LeuleuCysProProProArxLeuCyethrasphangilylImetIleaitrPaangly	220
Db		1385	TTGTGTGTGCTCTCCACAGCATATGACACTATATGACATTATGATTCATTCGA-----	14354
Oy		221	IleGlualgleuArgAlaglyleuengllylleleuHisAspiIleugluglylleargtyrlu	240
Db		1435	-----	14354
Oy		241	ProlyscysProleuGlyValaspIleserlysgluValIGlygluaIaserilleyVal	260
Db		1436	--TGATGTCTCTTTGAGTAGACATATCAAAAGAAAGTTGGAGAAAGCTTCATATAAGTA	14924
Oy		261	ProGlnleuLyMetGluIle	267
Db		1493	CCACAATTAAAAATGAGAGATA	1513
RESULT 13				
LOCUS	BC078974	1546 bp	mRNA	linear ROD 03-AUG-2004
DEFINITION	Rattus norvegicus cDNA clone IMAGE:711906, partial cds.			
VERSION	BC078974			
KEYWORDS	BC078974.1 GI:50926879			
SOURCE	.			
ORGANISM	Rattus norvegicus (Norway rat)			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
	Rattus.			
REFERENCE	1 (bases 1 to 1546)			
AUTHORS	Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shemen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.J., Heiseh, F., Ditchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stedilton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schwarz, T.E., Brownstein, M.J., Udwin, T.B., Toehiyuki, S., Caminci, P., Prange, C., Kana, S.S., Loquellano, N.A., Peters, G.U., Abramsom, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huyl, S.W., Willison, D.R., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, B., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bonfield, G.G., Blakesley, R.W., Touchman, J.W., Green, B.D.,			

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywnicki,M.I., Skalska,U., Smalins,D.E.,
Schmetch,A., Schein,J.F., Jones,S.J. and Warr,M.A. Initial
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932

2 (bases 1 to 1546)
Director MGC Project.
Direct Submission
Submitted (02-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cga@db-remail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILN)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.sngc.stanford.edu>
Contact: (Dickson, Mark) mcd@pacsl.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAX Plate: 182 Row: f Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers

1. 1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/tissue_type="Testis, rat (Brown Norway)"
/clone_id="NIH_MGC_237"
/lab_host="DH10B"
/note="vector: pExpress1"
144...>1546
/codon_start=1
/product="Unknown (protein for IMAGE:7111906)"
/protein_id="AAH78974.1"
/db_xref="GI:50926880"
/translation="MLMSKTKACAIIRPPRSANVRGFRRRVOPRALFHKLVLGIT
SCDTHAAVDDEGVNLDGEALHSQTEVHLTGIVIPVAOQLRENIQRIVEALSAS
GVSPSDLSAATNTIKGLALSLGVGSFSYVNOFKPEPIPIHHNEHALTLRLTHK
VGPPFLVILISGCHCLALAVOSVDFLLDKSIDIAGMDLDEVARSLIKIPECSY
MSGGKAIEHLAKGNPHFTINPMONAKKCDPSFGLOHTDXKLTTHKEKGKIG
OLISSADIIAAVOAHATAGHLARTRIALLFCOOKULSRANVAVSGVASNLVIR
RALEIRVANATQTLLCPPRPLCTDNIMIAMNGBIBRLPAELGLIHVEDIRPYKAGE
IVSMRLVATLTDLSVSSTHTVAHSPINSSGRGANIQFMCSCTGYMYRTVRYRTL
NNINERSKKKKKKKKK"

ORIGIN

Alignment Scores:

Pred. No.: 4.7e-100 Length: 1546
Score: 1116.00 Matches: 214
Percent Similarity: 92.56% Conservative: 10
Best local Similarity: 88.43% Mismatches: 18
Query Match: 80.58% Indels: 0
DB: 10 Gaps: 0

US-10-649-2_COPY_148_414 (1-267) x BC078974 (1-1546)

CY 1 MetGluAlahlgAlaleuTrllleArgyleuThAsnlyyVaIgIuhPeProPhelueVal 20
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 585 ATGAAGCGCTCAGCGGCTGACTATTAGGTCAACCCACAAATCGGANTTCCTTTTITTAGTT 644

QY 200 ThrleuEuCySPProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsn 219
DB 673 ACCTGTTGTCTCCACCTCCAGACTGTGCATGCAATGCAATGATTCATGAGAT 732
QY 220 GlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyr 239
DB 733 GGAATTGAAAGATTAAGTCTGCTGGCTGGGCGCTTTTACATGATGTAGAAAGACATCCGATAT 792
QY 240 GluProLysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLys 259
DB 793 GAACCAAAATATCTCTTGGAGTAGACATATCCAGAGAGTTGAGAAAGCTCCATTAATA 852
QY 260 ValProGlnLeuLysMetGluIle 267
DB 853 GTACCCGATTAATAATGCACTT 876
RESULT 15
BX934991 1522 bp mRNA linear VRT 02-FEB-2004
LOCUS BX934991
DEFINITION Gallus gallus finished cDNA, clone CHEST189114.
ACCESSION BX934991
VERSION BX934991.1 GI:41635519
KEYWORDS Gallus gallus (chicken)
SOURCE Gallus gallus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1522)
Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Gardner, C., Chalk, S.E.,
Crounch, M.D.R., Davies, R.M., Francis, M.D., Graham, D.V.,
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,
Mblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,
Tickle, C. and Wilson, S.A.
Direct Submission
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: chickens@hms.unist.ac.uk
BBRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
sequencing project.
This sequence is from the
BBRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
from a library constructed by Elizabeth Boech. cDNA was prepared
from RNA extracted from muscle, normalised, and poly A-tailed.
ECORI-NotI cut cDNA was then ligated into the vector. Vector:
pBlueScript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
coli DH10B.
FEATURES
Source 1..1522
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST189114"
/clone_lib="CSGREN11"
/dev_stage="adult"
ORIGIN
Alignment Scores:
Pred. No.: 1,24e-82 Length: 1522
Score: 939.00 Matches: 176
Percent Similarity: 81.06% Conservative: 38
Best Local Similarity: 66.67% Mismatches: 50
Query Match: 67.80% Indels: 0
DB: 5 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x BX934991 (1-1522)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 587 ATGAGGCTACGCACTTACATGACATGACAGCAAGTAGAATTTCCCTTCTAGTT 646
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

DB 647 CTTTAACTCTCCGAGAGTCACTGATCTTGGAGCTAGACAGGAGATTTCCTT 706
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaGlyArg 60
DB 707 CTGCTTGGACAGCTCCATTAATATAGCACAGGAGTGAATGCTGTAGAGTAGCAAGAG 766
QY 61 LeuSerLeuIleLysHisProGluCySPProLysSerThrMetSerGlyGlyLysAlaIleGluHis 80
DB 767 CTCTCTTTAGTAGACACCCGAGAGTCCACGGAGTGGCTGGGGGAGAGGCAATAGAGCAC 826
QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisAlaLys 100
DB 827 CTGGGCTCAAAACGAGAGCTGGAGACAGTACACTTGCACCTCCATGCAAGACTATCGT 886
QY 101 AsnCySPProPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 887 AACTGTGATTTTCTTTCTCCGAGCTTCAGAGCTTCGCAACAAACCAATCTTCAGAAA 946
QY 121 GluLysGluGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 947 GAAAGAGAGAGAGTATTAAGAGAGGAGAAATCTGTCTGCTTAAAGACATCGCTGCT 1006
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 1007 GCTGCACAGCAGTAGTGTGCTCTCATATTATTCACGGACACACCGACCATGCTTTC 1066
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
DB 1067 TGCATGAAAGACAGCATATTATTACCAAAAGCTGCAACTGTGGTGTATCAGAGAGATT 1126
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1127 GCAAGTAAACAGTATATATCAAGAAAGAGCTGCAAGCTTCGCAAAATCGTTTTCCT 1186
QY 201 LeuLeuCySPProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB 1187 TTTCTGTCTCTCTCCACAGCTGTGCACCGATATGATGTATGATTTGATGAAATGCG 1246
QY 221 IleguArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
DB 1247 ATTTGAAAGGTTCCGAGAGAGTGTGATTTTATACAGTACAGTATCCGCTTCGAA 1306
QY 241 ProLysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
DB 1307 CCAAAAGCTCCCTTGAATGATATTTCCAAAGAGTTGAAGAGATTCATCAAAAGTG 1366
QY 261 ProGlnLeuLys 264
DB 1367 CCAGACTTAAGG 1378

Search completed: February 16, 2005, 18:02:51
Job time : 4263.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:09:32 / Search time 508.461 Seconds
(without alignments)
3108.540 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALTRLRITNKVEPFLV.....DISKVEGASIKVPLQKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 segs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+g2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US10649273/runat_14022005_114702_16389/app_query.fasta_1.1429
-DB=n.Geneseq_16Dec04 -OPMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -POCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10649273 @CGN_1_1_1057 @runat_14022005_114702_16389 -ICPU=6 -ICPU=3
-NO_MMAP -LARGOUDERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:
1: geneseqn1980s:
2: geneseqn1990s:
3: geneseqn2000s:
4: geneseqn2001as:
5: geneseqn2001bs:
6: geneseqn2002as:
7: geneseqn2002bs:
8: geneseqn2003as:
9: geneseqn2003bs:
10: geneseqn2003cs:
11: geneseqn2003ds:
12: geneseqn2004as:
13: geneseqn2004bs:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1385	100.0	1416	8	ABX70950	Abx70950 Novel hum
2	1385	100.0	1526	6	ABs76635	Abas76639 DNA encod
3	1385	100.0	2058	6	ABA93268	Abas93268 Human O-s
4	1385	100.0	2197	6	ABs76635	Abas76665 DNA encod
5	1385	100.0	2572	8	ABt23207	Abt23207 Human pro

6	1348	97.3	1820	6	AAD46856	Aad46856 Human gly
7	1348	97.3	1820	8	ACA60887	Aca60887 Human cdn
8	1348	97.3	1821	10	ABs57020	Abas57020 CDNA enco
9	1204	86.9	2208	10	ADA52832	Ada52832 Human cod
10	1204	86.9	2890	12	ADQ24627	Adq24627 Human bot
11	995.5	71.9	3358	10	ADB31345	Adbs31345 Human dia
12	983	71.0	1572	6	ABQ75508	Abq75508 Murine si
13	599	43.2	2734	5	ASB84622	Asb84622 DNA encod
14	468	33.8	371	12	ADL66725	Adl66725 DNA up-re
15	468	33.8	371	12	ADL66726	Adl66726 DNA up-re
16	415.5	30.0	1601	4	ABL24633	AbL24633 Drosophi1
17	415.5	30.0	3656	4	ABL24632	AbL24632 Drosophi1
18	401.5	29.0	1557	3	ACA38454	Aac38454 Arabidops
19	342	24.7	4360	6	AAH48239	Aah48239 Ehrlichia
20	338.5	24.4	1385	4	AAH15110	Aah15110 Human cdn
21	321.5	23.2	1146	8	ACA26804	Aca26804 Prokaryot
22	314.5	22.7	936	8	ACA20445	Aca20445 Prokaryot
23	308.5	22.3	1044	8	ACA39102	Aca39102 Prokaryot
24	308.5	22.3	94750	4	AAF28551	Aaf28551 Genomic f
25	307	22.2	1053	12	ADL03120	Adl03120 DNA encod
26	304.5	22.0	1032	8	ACA43173	Aca43173 Prokaryot
27	301.5	21.8	1000	4	AAE91424	Aae91424 Moraxella
28	301.5	21.8	1000	6	ABK37804	Abk37804 DNA sequ
29	300	21.7	1092	6	ABQ90383	Abq90383 M. capaul
30	291.5	21.0	1026	4	AA554064	Aas54064 Pseudomon
31	291.5	21.0	1026	4	ACA42146	Aca42146 Prokaryot
32	291.5	21.0	1026	10	ADG73341	Adg73341 P aerugin
33	291.5	21.0	1059	11	ABD02280	Abd02280 Pseudomon
34	291.5	21.0	1206	11	ABD02197	Abd02197 Pseudomon
35	289.5	20.9	1026	10	ADG73343	Adg73343 P aerugin
36	289.5	20.9	1029	4	AA553309	Aas53309 Haemophil
37	289.5	20.9	1029	8	ACA34150	Aca34150 Haemophil
38	289.5	20.9	110000	2	AAT42063_05	Contamination (6 of
39	285	20.6	372	5	ABV61069	Abv61069 Human pro
40	282.5	20.4	9667	13	ADT05493	Adt05493 Haemophil
41	282.5	20.4	85814	13	ADT05464	Adt05464 Haemophil
42	280.5	20.3	1020	8	ACA44384	Aca44384 Prokaryot
43	280.5	20.3	1074	10	ADP02056	Adp02056 Bacterial
44	279.5	20.2	1014	4	AA556045	Aas56045 Salmonell
45	279.5	20.2	1014	8	ACA51431	Aca51431 Prokaryot

ALIGNMENTS

RESULT 1	
ID	ABX70950
ABX70950	standard; cDNA; 1416 BP.
XX	
AC	ABX70950;
XX	
DT	05-MAR-2003 (first entry)
XX	
DR	Novel human cDNA sequence #175.
XX	
KW	Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW	Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW	insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
KW	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW	Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW	differentiation; stem cell growth factor; haematopoiesis; chemokine;
KW	haemostatic; antiinflammatory; expressed sequence tag; ESR.
OS	Homo sapiens.
XX	
PN	W0200281731-A2.
XX	
PD	17-OCT-2002.
XX	
PF	29-JAN-2002; 2002WO-US001222.

XX 30-JAN-2001; 2001US-00774528.
 PR (HYSE-) HYSEQ INC.
 PA (GOOD) GOODRICH R W.
 XX
 XX Tang Y, Liu C, Zhou P, Asundi V, Zhang J, Zhao Q, Ren F,
 PI Xue AJ, Yang Y, Weinman T, Wang J, Wang D, Drmanac R;
 XX WPI; 2003-058563/05.
 DR
 XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic
 PT disorders, coagulation disorders, and inflammatory diseases.
 XX
 PS Claim 1; Page: 612pp; English.
 XX
 CC This invention relates to the cDNA sequences encoding an isolated novel
 CC human polypeptide. The protein encoded by the nucleic acid of the
 CC invention is useful for treating central and peripheral nervous system
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
 CC bacterial, viral or fungal infections; allergic conditions such as
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,
 CC infection or function of infectious agents such as bacteria, fungi,
 CC viruses, or to effect bodily characteristics, biorthylms or circadian
 CC cycles of rhythms. The protein may also have
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis
 CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
 CC activities. The cDNA sequences of the invention are useful for expressing
 CC recombinant protein for analysis. The present sequence represents a novel
 CC human cDNA sequence of the invention, this sequence is an expressed
 CC sequence tag (EST) and was identified using subtractive hybridisation
 XX
 SQ Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,68e-147 Length: 1416
 Score: 1385.00 Matches: 267
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x ABX70950 (1-1416)

QY 1 MetGUAHAHSALEuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
 DB 502 ATGAGGGCTCAGCACTTAATTAGGTGACCAATTAAGTAATTCCTTTTACGTT 561

QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
 DB 562 CTTTGGATTTCTGGAGGTCACCTGCTGTGGCATTTAGTTCAAGAGCTTTCAGATTTCTG 621

QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60
 DB 622 CTTCTTGGAAAGCTCTTGGACATACCAAGGTGACATCTTGAACAAGTGCAGAGAA 681

QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
 DB 682 CTTTCTTAAATAAACATCCAGAGGCTCCACCAAGAGGGTGGGAAAGCCATAGAACAT 741

QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLys 100

DB 742 TTGGCCAAACAGAAATATGATTTCATTTTGACATCAACCTCCCTTGATCATGCTAA 801
 QY 101 AenCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
 DB 802 AATTGGATTTCTTTTACTGACTCAACGTTACTGATTAATAATTAAGAAAG 861

QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
 DB 862 GAAAAAGGAAAGGATTTAGAAAGGGCAATCTCTCTTCAGCAGACATGCTGCC 921

QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
 DB 922 ACAGTACAGACACATATGATGATGATCTTGTGAAAGAAACATCGGGCTATTCTGTT 981

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
 DB 982 TGTAAAGCAGAGACATGTTTACCTCAAAATTAATGAGTACGTGTGCATCTGTGTC 1041

QY 181 AlaSerAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
 DB 1042 GCAAGTAACTTCTATATCCGACAGCTCTGAAATTTTAACAAACGACACAGTGCAT 1101

QY 201 LeuLeuCyAspProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
 DB 1102 TTGTTGTGCTCTCCCTCCAGACTATGCACTGATTAATGCAATTATGATTCATGGAATGCT 1161

QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
 DB 1162 AATTGAAAGCTAGCTGCTGCTGGCATTTTACATGATGAAAGGCATTCGCTATGAA 1221

QY 241 ProLysCyAspProLeuGlyValAspIleSerLysGlnValGlnValAspIleLysVal 260
 DB 1222 CCAAAATGCTCTCTTGGATGACATATCAAAAGATTGGAGAACTTCATTAAGTA 1281

QY 261 ProGlnLeuLysMetGluIle 267
 DB 1282 CCACATTAAATATGAGATA 1302

DB 1282 CCACATTAAATATGAGATA 1302

RESULT 2
 ABS76639
 ID ABS76639 standard; DNA; 1526 BP.
 XX
 AC ABS76639;
 XX
 DT 11-DEC-2002 (first entry)

DE DNA encoding novel human metalloprotease MP1 fragment #1.
 DE
 DE Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 KW neurological disorder; gene; de.
 KW
 XX Homo sapiens.
 OS
 XX
 XX MO200272751-A2.
 PN
 XX
 PD 19-SEP-2002.
 XX
 PF 05-FEB-2002; 2002MO-US003353.
 XX
 PR 05-FEB-2001; 2001US-0266518P.
 PR 10-APR-2001; 2001US-0282814P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

XX WPI: 2002-723329/78.
 DR P-PSDB; ABG96487.
 XX
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
 PT treating, or ameliorating diseases associated with aberrant
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
 PT neurological disorders.
 XX
 PS Disclosure: Page 462-463; 473pp; English.
 XX
 CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloproteinase (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This sequence represents a
 CC metalloproteinase MP1 polynucleotide
 CC
 SQ Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.2e-147 Length: 1526
 Score: 1385.00 Matches: 267
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x ABG76639 (1-1526)

QY 1 MetGuaIaAh;SAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
 DB 1 ATGAGGCTCCTGACCTTACTTATGAGTTGACCAATTAAGTGAATTTCTTTTATGTT 60
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
 DB 61 CTTTGTGATTTCTGAGGTCACGTGCTGTTGGCATTAGTTCAAGAGTTTCAATTTTCG 120
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
 DB 121 CTTCTTGAGAAAGTCTTGTGACATGACACCAAGGTGACATCTTGACAAAGGTGCAAGANA 180
 QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysValAlaIleGlnHis 80
 DB 181 CTTTCTTAAATTAACATCTGAGTCTCCACCAATGAGTGGAGAAACCATTAAGACT 240
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
 DB 241 TTGGCCAAACAAAGAAATGATTTCAATTTTGACATCAAACTCCCTTGATCATCTAA 300
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
 DB 301 AATGTGATTTTCTTACTGAGCTTCAACAGTTACTGATTAATAATATGAAG 360
 QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
 DB 361 GAAAAAGGAGAGATGATGAGAAAGGCGCAATTCCTTGACGACAAACATTCCTGCC 420
 QY 141 ThrValGlnHisHisMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
 DB 421 ACAGTACGACACAAATGCAATGCAATCTTGTAAGAAAGAACATCGGGCTATTCGTTT 480

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal 180
 DB 481 TGTAAAGCAGAGAGACTTGTACTCTCAAAATTAATGACGATGCTGTCATCTGCTGTC 540
 QY 181 AlaSerAspPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
 DB 541 GCAAGTAATCTTATATCCGACAGCTCTGGAATTTTAAACCAACACACAGTCACCT 600
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPAAsnGly 220
 DB 601 TGTGTGTGCTCTCTCCACAGCTATGCACTGATTAAGCAATTAATGATGATGAATGG 660
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyGlu 240
 DB 661 ATTGAAGACTAGCTGCTGCTGGCATTATTAACAGCATTAAGGCAATCCGCTATGA 720
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
 DB 721 CCAAAATGCTCTTGGAGTAGACATATCAAAAAGATTGAGAGAACTTCCATTAAGA 780
 QY 261 ProGlnLeuLysMetGluIle 267
 DB 781 CCACAAATTAATAATGAGATA 801
 RESULT 3
 ID ABA93268 standard; cDNA, 2058 BP.
 AC ABA93268;
 XX
 DT 19-APR-2002 (first entry)
 DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.
 KW Human; O-sialoglycoproteinase-like protein; OSGP, enzyme; gene; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 110..1354
 FT /tag= a
 FT /product= "O-sialoglycoproteinase-like protein"
 PN CN1318550-A.
 PD 24-OCT-2001.
 XX
 PF 19-APR-2000; 2000CN-00106834.
 PR 19-APR-2000; 2000CN-00106834.
 PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
 PI Mao Y, Xie Y;
 XX
 DR WPI: 2002-115090/16.
 DR P-PSDB; ABB05481.
 PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
 PT for diagnosing, preventing and treating related diseases.
 XX
 PS Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.
 XX
 CC The present sequence encodes human O-sialoglycoproteinase-like protein
 CC (OSGP). The present invention also describes: (1) the preparation of
 CC the OSGP protein; (2) applying the OSGP protein in diagnosis; (3) the
 CC prevention and/or treatment of related diseases; (4) utilizing the OSGP
 CC protein in screening its agonist, excitomotor and inhibitor and preparing
 CC an antibody against the OSGP protein; and (5) the use of the OSGP
 CC polynucleotide sequences, proteins, agonists, excitomors, inhibitors
 CC and antibodies in treating diseases related to the abnormal OSGP gene
 CC and in preparing the medicine composite for the treatment
 XX

Db 672 ATGAGGCTCAGCAGCTTACTTATTAGTTGACCAATTAAGTAATTTCTTTTATGTT 731
Qy 21 LeuLeuLeuSerGlyGlyVhiCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 732 CTTTGTATTCTGGAGGTCACGTCTGTGGCATTTAGTTCAAGAGGTTTCAATTTTCTG 791
Qy 41 LeuLeuGlyLysSerLeuAspL1eAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 792 CTTTGTGAAAGTCTTTGACATGACACAGGTGACATCTTTGACAGGTGCAAGAGA 851
Qy 61 LeuSerLeuL1eLysH1eProGlyCysSerThrMetSerGlyLysValAlaL1eGlnHis 80
Db 852 CTTTCTTAATTAATAACATCCAGAGTGTCTCCACATGAGGTGGGAAAGCCATAGAACAT 911
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspL1eLysProProLeuHisHisAlaLys 100
Db 912 TTGGCCAAACAGAAATGATTTGATTTGACATCAACCTCCCTTGATCATCTTAA 971
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysL1eL1eMetLysLys 120
Db 972 AATTGTATTTTCTTTACTGAGCTTCAACGCTTACTGATTAATTAATTAAGAAAAG 1031
Qy 121 GlnLysGlnGlnGlyL1eGlnLysGlyL1eLeuSerSerAlaAlaAspL1eAlaAla 140
Db 1032 GAAAAAGAGAAAGGATTGAGAGAGGCAAAATCCTGCTTGAGCAGCAGCATTTGCTGCC 1091
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgHisArgAlaL1eLeuPhe 160
Db 1092 ACAGTACAGCAGACAAATGCGATGTCATCTTGAAAAAACAACATCGGCTTATCTGTTT 1151
Qy 161 CysLysGlnAspArgPheLeuProGlnAsnAsnAlaValL1eLeuValAlaSerGlyLysVal 180
Db 1152 TGTAAAGCAGAGACTTGTACTCTCAAAATTAATGACATGCTGTGCAATCTGTGTGTC 1211
Qy 181 AlaSerAsnPheTyrL1eArgArgAlaLeuGlnL1eLeuThrAsnAlaThrGlnCysThr 200
Db 1212 GCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTAAACAAACGACACAGTGCAT 1271
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyL1eMetL1eAlaTrpAsnGly 220
Db 1272 TTGTTGTCTCTCTCCACAGCATATGACATTAATGACATTAATGACATGAATGCT 1331
Qy 221 L1eGlnArgLeuArgAlaGlyL1eGlnGlyL1eLeuHisAspL1eGlnGlyL1eArgTrpGlu 240
Db 1332 ATTGAAAGACTACGCTGCTGGCGATTTTACATGACATGAAGAGGACATCGCATATGAA 1391
Qy 241 ProLysCysProLeuGlyValAspL1eSerLysGluValGlyGlnAlaSerL1eLysVal 260
Db 1392 CCAAAATGCTCTCTTGAGTGACATATCAAAAGAAAGTTGAGAAAGCTTCCATTAAGTA 1451
Qy 261 ProGlnLeuLysMetGlnL1e 267
Db 1452 CCACAAATTAATAATGAGATA 1472
RESULT 5
ABT23207
ID ABT23207 standard; DNA: 2572 BP.
XX AC ABT23207;
XX
XX 01-MAY-2003 (first entry)
XX
XX Human protein modification + maintenance molecule DNA SEQ ID No 36.
XX
XX Cerebrostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
XX cerebroprotective; antiparinsonian; nootropic; antiinflammatory;
XX antitumor; hepatotropic; gynecological; antibacterial; virucide;
XX prototoxoid; antiparasitic; cell proliferative disease; PMOD;
XX protein modification and maintenance molecule; immunogenic fragment;
XX cancer; autoimmune; inflammatory disease; neurological disorder;
XX gastrointestinal; developmental; vesicle trafficking disorder; infection;
XX protein-protein interaction; drug-target interaction;
XX gene expression profile; human; gene; ds.

XX OS Homo sapiens.
XX XX
XX PN WO2003000844-A2.
XX
XX PD 03-JAN-2003.
XX
XX PF 18-JUN-2002; 2002WO-US019360.
XX
XX PR 22-JUN-2001; 2001US-0300508P.
XX PR 06-JUL-2001; 2001US-0303445P.
XX PR 13-JUL-2001; 2001US-0305405P.
XX PR 09-AUG-2001; 2001US-0311442P.
XX PR 24-AUG-2001; 2001US-0314821P.
XX PR 29-AUG-2001; 2001US-0315992P.
XX PR 03-MAY-2002; 2002US-0378205P.
XX
XX PA (INCY-) INCYTE GENOMICS INC.
XX
XX PI Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM,
XX Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H,
XX Forgythe J, Barroso I, Ramkumar J, Griffin JA, Li UX, Yang J,
XX Thangaveju K, Gietzen KU, Ding L, Baughn MR, Borowsky ML, Yao MG,
XX Walla NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK,
XX Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebartjadian Y;
XX WPI; 2003-184039/18.
XX DR P-PSDB; AB026654.
XX
XX PT New isolated human PMOD polypeptide and polynucleotide, useful for
XX diagnosing, treating and preventing diseases or conditions associated
XX with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
XX infections.
XX
XX PS
XX
XX SS Claim 91; Page 211; 225pp; English.
XX
XX CC The invention relates to an isolated polypeptide comprising: any of 28
XX sequences of 48-1256 amino acids; a natural amino acid sequence at least
XX 90% identical to the 28 amino acid sequences; 94% identical to a sequence
XX of 703 or 267 amino acids; 96% identical to a sequence of 414 amino
XX acids, or 97% identical to a sequence of 242 amino acids, all given in
XX the specification; or a biologically active or immunogenic fragment of
XX the isolated polypeptide. The polypeptides and polynucleotides are useful
XX in diagnosing, treating and preventing diseases or conditions associated
XX with the decreased expression of protein modification and maintenance
XX molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
XX atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
XX allergies), neurological disorders (e.g. stroke, Parkinson's disease,
XX epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
XX endometriosis), developmental, vesicle trafficking disorders, and
XX infections (e.g. bacterial, viral, parasitic, protozoal). These are also
XX useful in assessing the effects of exogenous compounds on the expression
XX of nucleic acid and amino acid sequences of PMOD. The PMOD or its
XX fragments are useful in screening compounds for effectiveness as agonist
XX or antagonist of the polypeptides, or in altering the expression of the
XX target polynucleotide and compounds that specifically bind to or modulate
XX the activity of the polypeptide. The microarray is useful in monitoring
XX CC or measuring protein-protein interactions, drug-target interactions, and
XX CC gene expression profiles. This polynucleotide sequence represents the DNA
XX encoding a human PMOD protein of the invention
XX
XX SQ Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1,096-146
XX Score: 1385.00
XX Percent Similarity: 100.00%
XX Best Local Similarity: 100.00%
XX Query Match: 100.00%
XX DB: 8
XX Gaps: 0
XX
XX US-10-649-273-2_COPY_148_414 (1-267) x ABT23207 (1-2572)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
 DB 585 ATGAGGCTCATGCACTTACTATTAGTTGACCAATAAGTAATTCCTTTTAACT 644
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
 DB 645 CTTTGTGATTTCTGGAGGCTCATGCTGTGGTCATTAGTTCAAGGAGTTTCAAGATTTTCTG 704
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
 DB 705 CTTCTGGAAAGCTCTTTGGACATAGCACCGAGTGACATGCTTGACAGAGTGCGAAGAA 764
 QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
 DB 765 CTTTCTTTAATAAACAATCCAGAGTGCTCCACATAGAGTGAGGAAAGCCATTAGAACT 824
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
 DB 825 TTGGCCAAACCAAGAAATAGATTTCATTTTGACATCAAACTCCCTTGCATCATCTAAA 884
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
 DB 885 AATTGTGATTTTCTTTACTGAGCTTCAACAGTTACTGATAAATAATTAATGAAGAAAG 944
 QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
 DB 945 GAAAAGAGGAAGATTAAGAGAGGGCAATCCTGCTCTTCAGCAGCAGACATTCCTGCC 1004
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
 DB 1005 ACAGTACAGCACACAATGAGCATGTCTGCTGTGAAAAGAACATCGGGCTATTCTGTTT 1064
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValAlaLeuValAlaSerGlyGlyVal 180
 DB 1065 TGTAAAGCAGAGACTGTTTACTCAATAATATGACAGTCTGTTGCACTGTGGTGTGC 1124
 QY 181 AlaSerAsnPheTYR11leArgArgAlaLeuGlnIleLeuThrAspAlaThrGlnCysThr 200
 DB 1125 GCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTAAACAAGCAACAGCTGCACT 1184
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
 DB 1185 TTGTTGTCTCCCTCCAGACTATGACATGATATGCAATTATGATGATGATGATGAT 1244
 QY 221 IleGluArgLeuArgAlaGlyLeuGlnIleLeuHisAspIleGlnGlyIleArgTYR12 240
 DB 1245 ATGTAAACACATACGCTGGCTGTGGCATTTTACATGACTTGAAAGGACATCCGCTATGA 1304
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyAlaSerIleLysVal 260
 DB 1305 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTTGAGAAAGCTTCCATTAAAGTA 1364
 QY 261 ProGlnLeuLysMetGluIle 267
 DB 1365 CCACAATTAAAAATGAGATA 1385

RESULT 6

AAID6856
 ID AAD46856 standard; cDNA, 1820 BP.

AC AAD46856;

DT 27-JAN-2003 (first entry)

DE Human glycoprotease 28472 cDNA.

KM Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
 KM 7TM; glycoprotease; immune disorder; Iga deficiency; allergy; arthritism;
 KM rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;
 KM hypertension; ischemic heart disease; obesity; myocardial infarction;
 KM endothelial cell disorder; Grave's disease; psoriasis; brain disorder;
 KM Parkinson's disease; Alzheimer's disease; haematopoietic disorder;
 KM cerebral oedema; metabolic disorder; liver disorder; platelet disorder;

KM chromosome mapping; tissue typing; gene therapy; neuroprotective;
 KM cytosolic; anorectic; cardiac; haemostatic; gene; ss.
 XX Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 146..1390
 FT /tag= a
 FT /product= "Human 28472 protein"
 FT /note= "This region is specifically claimed as SEQ ID NO:
 FT 6 in claim 1 of the specification"

WO200274960-A2.

26-SEP-2002.

08-NOV-2001; 2001WO-US051427.

08-NOV-2000; 2000US-0246768P.

08-NOV-2000; 2000US-0246772P.

15-NOV-2000; 2000US-0249185P.

(MILL-) MILLENNium PHARM INC.

Leiby KR, Kapeller-Liebermann R, Glucksmann M;

WPI: 2002-759898/82.

P-PSDB; AAE29234.

New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
 useful for diagnosing and treating cancer, immune, cardiovascular,
 hemopoietic, brain, pain, metabolic, liver or platelet disorders, and
 in pharmacogenomics.

Claim 1; Fig 8; 178pp; English.

The present invention relates to novel 38650, 28472, 5495, 65507, 81588
 or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
 protease or seven transmembrane domain (7TM) receptor family members.
 Sequences of the invention are useful in diagnosing and treating cancer
 or aberrant cellular proliferation and/or differentiation (e.g. colon or
 lung cancer), immune disorders (e.g. selective Iga deficiency, rheumatoid
 arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
 hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
 myocardial infarction, thrombus) including endothelial cell disorders
 (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
 disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
 pain and metabolic disorders (e.g. obesity), liver disorders or platelet
 disorders. They are also useful in screening assays, predictive medicine
 (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
 and pharmacogenetics) and prophylactic and therapeutic methods. The
 and nucleic acids may also be used in chromosome mapping, tissue typing and
 forensic biology and as surrogate markers. Sequences of the invention are
 also used in gene therapy. The present sequence is human glycoprotease
 28472 cDNA

Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;

Alignment Scores:

Prod. No.: 1,07e-142 Length: 1820
 Score: 1348.00 Matches: 260
 Percent Similarity: 98.50% Conservative: 3
 Best Local Similarity: 97.38% Mismatches: 4
 Query Match: 97.33% Indels: 0
 DB: 6 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x AAD46856 (1-1820)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
 DB 585 ATGAGGCTCATGCACTTACTATTAGTTGACCAATAAGTAATTCCTTTTAACT 646
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

Db	647	CTTTTGATTTCCTGGAGTCACCTGCTCTGTTGGCATTAAGTTCAAGAGCTTCAGATTTTCTG	706
QY	41	LeuLeuGlyLYSserLeuaspIlealaProGlyaAspMetLeuaspLYsValalaArg	60
Db	707	CTTCTGGAAAGTCTTTGGACATACACCAAGGTGATCTCTTGACAAAGGTGGCAAGA	766
QY	61	LeuSerIleuIleLYHisProGluCYsSerThrMetSerGlyLYsAlaIleGluHis	80
Db	767	CTTTCTTTAAATAAACATCCACAGAGCTCCACCATGAGTGGGGAAGGCCATTAACAT	826
QY	81	LeuAlaLYSGInGlyAsnArgPheHisPheaspIleLYsProProLeuHisIleAlaLYs	100
Db	827	TTGGCCAAACAGAGAAATRGATTTTCATTTTGACATCAACCTCCCTTGCAATCTTAA	886
QY	101	hencYasPpheSerPheThrGlyLeuGlnHisValThrAspLYsIleIleMetLYsLYs	120
Db	887	AATGTGATTTTCTTTTACTGCACTTCAACGCTTACTGATTAATAATGAAACAG	946
QY	121	GluLYSGInGluGlyIleGlyLYsGlyGlnIleLeuSerSerAlaIleaspIleAla	140
Db	947	AAACAGAGGAGGATGTAGAGAAAGGGCAATCCGTCTTCAGCAGACATTCCTCC	1006
QY	141	ThrValGlnHisIleThrMetAlaCYHisIleuValLYsArgThrHisArgAlaIleuPhe	160
Db	1007	ACAGTACGACACACATGCGATGTCATCTTGTAAGAAACACATCGGGCTATTTGTTT	1066
QY	161	CysLYSGInArgAspLeuLeuProGlnAsnAlaValaLeuValaIleSerGlyLYsVal	180
Db	1067	TGTAAAGCAGAGAGACTTGTTCCTCAAAATAATGACATGCTGTGCATCTGGTGTCTC	1126
QY	181	AlaSerAsnPheTYrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr	200
Db	1127	GCAAGTAACCTCTAATATCCGCGACAGCTCTGGAAATTTTAACAAAGCAACAGTGCAC	1186
QY	201	LeuLeuCYsProProProArgLeuCYsThrAspAsnGlyIleMetIleAlaThrAsnGly	220
Db	1187	TTGTTGTCTCCCTCCGCCAGACTAAGCATGTAAATGCAATTAATGCAATGCAATGAT	1246
QY	221	IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTYrGlu	240
Db	1247	ATTGAAAGACATRGCGTCTGCTGGGCACTTTTACATGACATRGAGGCAATCCGCTATA	1306
QY	241	ProLYsCYsProLeuGlyValaAspIleSerLYSGInValaGlyGluAlaSerIleLYsVal	260
Db	1307	CCAAATGTCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCATTAAGTA	1366
QY	261	ProGlnLeuLYsMetGluIle 267	
Db	1367	CCACAATTAAATAATGAGATA 1387	
RESULT 7			
ACA60887			
ID	ACA60887	standard; cDNA; 1820 BP.	
AC	ACA60887;		
XX			
DT	08-JUL-2003	(first entry)	
XX			
DE	Human CDNA 28472	encoding a glycoprotease.	
XX			
KM	Human; ss; gene; cancer; aberrant cellular proliferation;		
KM	differentiation; immune disorders; heart disorder; brain disorder;		
KM	cardiovascular disorder; endothelial cell disorder; pain disorder;		
KM	haematopoietic disorder; blood vessel disorder; metabolic disorder;		
KM	liver disorder; platelet disorder; glycoprotease.		
OS	Homo sapiens.		
XX			
XX			
Key	Location/Qualifiers		
PT	146..1390		
PT	CDS		
PT	/tag= a		
PT	/product= "Glycoprotease"		

/note= "This CDS is specifically claimed in claim 1"

US2003009017-A1.

09-JAN-2003.

08-NOV-2001; 2001US-00012140.

08-NOV-2000; 2000US-0246768P.

08-NOV-2000; 2000US-0246772P.

15-NOV-2000; 2000US-0249185P.

(LEIB/) LEIBY K. R.
(KAE/) KAPELIER-LIBERMANN R.
(GLUC/) GLUCKSMANN M. A.

Leiby KR, Kapellier-libermann R, Glucksmann MA;
MPI. 2003-428888/40.
P-PSDB; ABU09569.

New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid molecules, useful for diagnosing, treating cancer, pain, or immune, heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic and liver disorders.

Claim 2, Fig 8; 90p; English.

The invention relates to an isolated 38650 (encoding adenosine deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7 transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or a sequence which is at least 60% identical to the six nucleic acids or their open reading frames, fragments of at least 15 nucleotides, CC naturally occurring variants, or a DNA insert of the plasmid deposited with the American Type Culture Collection as Accession No. not defined in the specification, which encodes the amino acid sequence). Also included are a host cell containing the nucleic acids (used to produce the protein), the encoded proteins, an antibody that selectively binds to the polypeptide, and identifying a compound that binds to/modulates the activity of the polypeptide. The nucleic acid molecules, polypeptides and methods are useful for diagnosing, treating cancer, aberrant cellular proliferation and/or differentiation, immune disorders, heart disorders, cardiovascular disorders including endothelial cell disorders, haematopoietic disorders, blood vessel disorders, brain disorders, pain and metabolic disorders, liver disorders and platelet disorders (many examples of these disorders are given in the specification). The present sequence is the human cDNA 28472 encoding a glycoprotease

Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	1,07e-142	Length:	1820
Score:	1348.00	Matches:	260
Percent Similarity:	98.50%	Conservative:	3
Best Local Similarity:	97.38%	Mismatches:	4
Query Match:	97.33%	Indels:	0
DB:	8	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x ACA60887 (1-1820)

1 MetGuaAaHsAlaLeuThrIleArgLeuThrAsnLysValGlnuPheProPheLeuVal 20
587 ATGAGGGCTCATGCACTTACTATTAGGTGACCAATAAGTAGAATTCTTTTATGTT 646
21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
647 CTTTGAATTTCTGAGGCTCAGCTCTGTGGCATTTAGTTCCAGAGATTTCTG 706
41 LeuLeuGlyIleValSerIleuAspIleAlaProGlyAspPheLeuAspLysValAlaArgArg 60
707 CTTCTTGGAAGACTCTTTGAGCATAGACCAAGGATGACATCTTACCAAGGGGCAAGAGA 766
61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80


```

Db      767 CTTCTTTAATAAATCCAGAGTCTCCACATGAGTGGGAAAGCCATAGACAT 826
Qy      81 LeuAlaIysGInGlyAAspArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db      827 TTGGCCAAACAGAAATAGATTTTCATTTTATGACATCAACCTCCCTTCATCATATCTAA 886
Qy      101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      887 AATTGTGATTTTCTTTTACTGACCTTCAACACGTTACTGATTAATAATATCAAAACAG 946
Qy      121 GlnLysGlnGlnGlyIleGlnLysGlnIleLeuSerSerAlaAlaAspIleAla 140
Db      947 AAACAAGAGAAAGATATGAGAGGGCAATCCGTCTTCAGCAGACAGATTCCTGCC 1006
Qy      141 ThrValGlnHisThrMetAlaCysHisLeuValLysAspGlnHisArgAlaIleLeuPhe 160
Db      1007 ACAGTACGACACACAAATGACATGTCATCTTGAAAAAAGAACACATCGGCTATTTCTGTTT 1066
Qy      161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyVal 180
Db      1067 TGTAAAGCAGAGACTTCTTACCTCAAAATATGACATGCTTGTCATCTGTCGTCTC 1126
Qy      181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
Db      1127 GCAAGTAACTTCTATATCCGACAGACTCTGGAAATTTTAAACCAACAGACAGTCACT 1186
Qy      201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
Db      1187 TTGTTGTGTCCTCTCCAGACTATGACCTGATTAATGACATTAATGACATGAAATGCT 1246
Qy      221 IleGlnArgLeuArgAlaGlyLeuGlnIleLeuHisAspIleGlnGlyIleArgTyrglu 240
Db      1247 ATTAAAGACTACACTGCTGCTGGCATTTTACATGACATTAAGAGGCAATCCGCTATGAA 1306
Qy      241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 260
Db      1307 CCAAAAGTCTCTTGGAGTACATCAAAAGAAAGTTGGAGAGCTTCCATAAAGTA 1366
Qy      261 ProGlnLeuLysMetGlnIle 267
Db      1367 CCACAATTTAAATAAGAGATA 1387

```

RESULT 8
ABSS7020
ID ABSS7020 standard; cDNA; 1821 BP.
XX
AC ABSS7020;
XX
DT 30-JAN-2003 (first entry)
XX
DE cDNA encoding novel human glycoprotease 28472.
XX
KW Cancer; aberrant cell proliferation; aberrant cell differentiation;
KW breast cancer; ovarian cancer; prostate cancer; colon cancer;
KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;
KW endothelial disorder; hematopoietic disorder; blood vessel disorder;
KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;
KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;
KW autoimmune disorder; hypertension; atherosclerosis; heart failure;
KW myocardial infarction; ischemic heart disease; Crohn's disease;
KW Graves' disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
KW cerebral ischemia; peripheral neuropathy; Alzheimer's disease;
KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT CDS 147..1391
FT /tag= a
FT /product= "Glycoprotease 28472"
FT /note= "Specifically claimed in claim 1"

```

PN      W0200277233-A2.
XX
PD      03-OCT-2002.
XX
PF      08-NOV-2001; 2001WO-US046724.
XX
PR      08-NOV-2000; 2000US-0246768P.
PR      08-NOV-2000; 2000US-0246772P.
PR      15-NOV-2000; 2000US-0249185P.
XX
PA      (MILL-) MILLENNIUM PHARM INC.
PI
PI      Leiby KR, Kapeller-Libermann R, Glucksmann M;
XX
XX      WPI: 2003-029938/02.
XX
DR      P-PSDB; ABG71162.
XX
PT      New adenosine deaminase, glycoprotease and seven transmembrane domain
PT      nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
PT      81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
PT      hypertension.
XX
PS      Claim 2; Fig 8A-B; 178pp; English.
XX
XX      The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
CC      14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
CC      38650 nucleic acid molecule comprises a sequence encoding adenosine
CC      deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
CC      a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
CC      sequences that encode a human seven transmembrane domain (7TM). The
CC      38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
CC      sequences are useful for diagnosing, preventing or treating a subject
CC      with or at risk of developing a disorder, e.g. cancer or aberrant
CC      cellular proliferation and/or differentiation (e.g. breast, ovarian,
CC      prostate, colon or lung cancer), immune disorders, heart disorders,
CC      cardiovascular disorders, endothelial disorders, hematopoietic disorders,
CC      blood vessel disorders, brain disorders, pain and metabolic disorders,
CC      liver disorders or platelet disorders. These disorders include carcinoma,
CC      sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
CC      hypertension, atherosclerosis, heart failure, myocardial infarction,
CC      ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
CC      syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral
CC      neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
CC      cachexia or diabetes. This sequence encodes the novel human glycoprotease
CC      28472
XX
SQ      Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;
XX

```

Alignment Scores:

Pred. No.:	1,076-142	Length:	1821
Score:	1348.00	Matches:	260
Percent Similarity:	98.50%	Conservative:	3
Best Local Similarity:	97.38%	Mismatches:	4
Query Match:	97.33%	Indels:	0
DB:	10	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x ABSS7020 (1-1821)

```

Qy      1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 20
Db      588 ATGAGAGGCTCATGCACTTACTATTAGTTGAGCCAAATTAAGTAAATTTCTTTTATGTT 647
Qy      21 LeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      648 CTTTGTGATTTCTGGAGGTCACGTCTGTTGGCATTAAGTTCACAGAGATTTTCAGATTTTCTG 707
Qy      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db      708 CTTCTTGAAGATCTTTTGACATAGACACAGGTGACATCTTGACAAAGTGGCAAGAA 767
Qy      61 LeuSerLeuIleLysHisPheProGlnCysSerThrMetSerGlyIleLysAlaIleGlnHis 80
Db      768 CTTTCTTTAATAAATCATCCAGAGTCTCCACCATGAGTGGTGGAAAGCCATAGAAAT 827

```

QY 81 LeuAlaIysGInGIYAsnArgPheHisPheAspIleIysProProLeuHisHisAlaIys 100
 DB 828 TTGGCCAAAGAGAGAAATGATTTGATTTGATCAATCAACCTCCCTTGATCATCTAA 887
 QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValIThrAspLysIleIleMetLysLys 120
 DB 888 AATGTGATTTTCTTTACTGACTGACTTCAACGCTTACTGATTAATAATGATAAACAAG 947
 QY 121 GlnIlySerIleGluGlyIleGlnIlySerIleLeuSerSerAlaIleAspIleAla 140
 DB 948 AAACAAAGAGAGAAATGATTTGATCAATCAACCTCCCTTGATCATCTAA 1007
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIlysaGThrHisArgAlaIleLeuPhe 160
 DB 1008 ACAGTACAGCACAATGAGCATGTCATCTTGTGAAAAGAACATCGGCTATTTGTTT 1067
 QY 161 CysIlySGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyGlyVal 180
 DB 1068 TGTAAAGCAGAGAGACTTGTACTTCAAAATATGACAGTACTGTCATCTGTGTC 1127
 QY 181 AlaSerAsnPheIleIleArgAlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThr 200
 DB 1128 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACGAAACGACATGCACT 1187
 QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly 220
 DB 1188 TTGTTGTGCTCTCCAGCATGACATGATTAATGACATTAATGACATGAGAAATGT 1247
 QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGly 240
 DB 1248 ATTGAAGAGCTACGTCGTGGCTTGGCATTTTACATGACATGAAGGATCCGCTATGA 1307
 QY 241 ProIlyCysProLeuGlyValAspIleSerIlySGlnValIleGlnAlaSerIleLysVal 260
 DB 1308 CCAAAAGTCTCTCTGAGTACATATCAAAAGAGTTGAGAGAGCTTCATAAAGTA 1367
 QY 261 ProGlnLeuIlyMetGlnIle 267
 DB 1368 CCACAAATTAATAATGAGATA 1388
 RESULT 9
 ADA52832
 ID ADA52832 standard; cDNA; 2208 BP.
 AC ADA52832;
 XX 20-NOV-2003 (first entry)
 DT XX Human coding sequence, SEQ ID 400.
 DE XX
 XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KM Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
 XX Homo sapiens.
 OS
 PN EP1293569-A2.
 XX
 PD 19-MAR-2003.
 XX
 PF 21-MAR-2002; 2002BP-00006586.
 XX
 PR 14-SBP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 PI Iesgai T, Sugiyama T, Otsuki F, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maehno Y,
 XX

DR WPI; 2003-395539/38.
 DR P-PSDB; ADA544471.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 PS
 PS Claim 1; SEQ ID NO 400; 2055pp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 XX
 SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,3e-126 Length: 2208
 Score: 1204.00 Matches: 239
 Percent Similarity: 89.51% Conservative: 0
 Best Local Similarity: 89.51% Mismatches: 4
 Query Match: 86.93% Indels: 24
 DB: 10 Gaps: 1
 US-10-649-273-2_COPY_148_414 (1-267) x ADA52832 (1-2208)
 QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValIleGlnPheProPheLeuVal 20
 DB 785 ATGGAGGCTCAGACACTTACTATTTAGTTGACCAATTAAGTAAATTTCTTTTACTT 844
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValIleGlnValSerAspPheLeu 40
 DB 845 CTTTGTATTTCTGAGAGCTACTGCTGTTGGATTAAGTTCAGAGAGCTTCAGATTTCTG 904
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
 DB 905 CTTCTTGGAAGACTCTTTGGACATGACACCGAGTGACATCTTGACAGAGTGGCAAGAA 964
 QY 61 LeuSerLeuIleLysHisPheProGlnCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
 DB 965 CTTCTTTAATTAATTAATCAATCAGAGTCTCCACATGAGTGGAGAAACCATTAAGCAT 1024
 QY 81 LeuAlaIysGInGIYAsnArgPheHisPheAspIleIysProProLeuHisHisAlaIys 100
 DB 1025 TTGGCCAAAGAGAGAAATGATTTGATTTGATCAATCAACCTCCCTTGATCATCTAA 1084
 QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValIThrAspLysIleIleMetLysLys 120
 DB 1085 AATGTGATTTTCTTTACTGACTGACTTCAACGCTTACTGATTAATAATGATAAACAAG 1144
 QY 121 GlnIlySerIleGluGlyIleGlnIlySerIleLeuSerSerAlaIleAspIleAla 140
 DB 1145 GAAAAGAGAGAGATTAAGAGAGGGAATCTGTCCTTGACAGCACAATGTCGCC 1204
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIlysaGThrHisArgAlaIleLeuPhe 160
 DB 1205 ACAGTACAGCACAATGAGCATGTCATCTTGTGAAAAGAACATCGGCTATTTGTTT 1264
 QY 161 CysIlySGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyGlyVal 180
 DB 1265 TGTAAAGCAGAGAGACTTGTACTTCAAAATATGACAGTACTGTCATCTGTGTC 1324
 QY 181 AlaSerAsnPheIleIleArgAlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThr 200
 DB 1325 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACGAAACGACATGTCACAT 1384
 QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly 220
 DB 1385 TTGTTGTGCTCTCCCTCCAGCATGACATGATTAATGACATTAATGATCA----- 1435
 QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGly 240

Db 1435 ----- 1435
QY 241 ProlyscysProleuGlyValaspilleserlysgluValgluAlaSerilleyVal 260
Db 1436 ---TGATGTCCTCTTGAGTACATATCAAAAGAGTTGAGAGCTTCCATAAAGTA 1492
QY 261 ProgluLeuLysMetGluIle 267
Db 1493 CCACATTTAAATAATGAGATTA 1513

RESULT 10
ADQ24627
ID ADQ24627 standard; DNA; 2890 BP.
XX
AC ADQ24627;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.
XX
KM soft tissue sarcoma; cytosol; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN MO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003MO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnick A;
XX
XX WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 7447; 210bp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytosolic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;

Alignment Scores:
Pred. No.: 4.83e-126 Length: 2890
Score: 1204.00 Matches: 239
Percent Similarity: 89.51% Conservative: 0
Best Local Similarity: 89.51% Mismatches: 4
Query Match: 86.93% Indels: 24
DB: 12 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x ADQ24627 (1-2890)

QY 1 MecGlulalHisAlaLeuThrIleArgLeuThrAenLysValGluPheProPheLeuVal 20

Db 1442 ATGAGGCGCTCATGCACTTACTACTTAGTTGACCAATAAGTAATTCCTTTTAGTT 1501
QY 21 LeuLeuIleSerGlyGlyPheCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 1502 CTTTGGATTTCTGGAGGTCACGTCTGTGGCATTAAGTTCAGAGAGCTTTCAGATTTTCG 1561
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 1562 CTTCTTGGAAAGTCTTTGGACATAGCACCAGGTGACATCTTGACAAAGTGGCAAGAA 1621
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlnHis 80
Db 1622 CTTCCCTTTAATAAATCATCAGAGTGTCTTCACATGAGTGTGGGAAACCATAGAAACAT 1681
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 1682 TTGGCCAAACAAAGAAATGATTTGATTTGACATCAAACTCCCTTGACATCATGCTTAA 1741
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 1742 AATGTGATTTCTTTTACTGACTTCAACAGCTTACTGATTAATAATTAATGAATAAG 1801
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db 1802 GAAAAAGAGAGAGATTTAGAGAGGGGCAATCTGTCTTCAGCAGACATGCTGCGC 1861
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 1862 ACAGTACAGCACAATGCGATGTCATCTTGAGAAAGAACATCGGGCTAATTCGTTT 1921
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
Db 1922 TGTAAAGCAGAGACTTGTACTCTCAAAATTAATGACGATCTGATGATCTGATGATGTC 1981
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 1982 GGAATTAATCTTGTATCCGACAGACTCGGAAATTTTAACAAACCAACACAGTGCATCT 2041
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIlePheAsnGly 220
Db 2042 TTGTTGTGTCCTCTCCACGATGACCTGATGACCTGATGACCTGATGACCTGATGACCT 2092
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrgLys 240
Db 2092 ----- 2092
QY 241 ProlyscysProleuGlyValaspilleserlysgluValgluAlaSerilleyVal 260
Db 2093 ---TGATGTCCTCTTGAGTACATATCAAAAGAGTTGAGAGCTTCCATAAAGTA 2149
QY 261 ProgluLeuLysMetGluIle 267
Db 2150 CCACATTTAAATAATGAGATTA 2170

RESULT 11
ADEB31345/C
ID ADEB31345 standard; DNA; 3358 BP.
XX
AC ADEB31345;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID NO 100.
XX
KW diagnostic and therapeutic polynucleotide; dithp; antihypertensive; antidiabetic;
KW antiinflammatory; cerebroprotective; antilipemic; antidiabetic;
KW immunosuppressive; neuroprotective; moodotropic; neuroleptic; tranquilizer;
KW osteopathic; antiallergic; antirheumatic; cytosolic; hepatotropic;
KW virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;
KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
KW thrombolytic; anticoagulant; anorectic; vasodilator; antitumor;
KW gene therapy; protein replacement therapy; human; gene; ds.

XX Homo sapiens.
 OS
 PN W02003062376-A2.
 XX
 PD 31-JUL-2003.
 XX
 PF 13-JAN-2003; 2003WO-US001096.
 XX
 PR 16-JAN-2002; 2002US-0349384P.
 PR 17-JAN-2002; 2002US-0349413P.
 PR 17-JAN-2002; 2002US-034946P.
 XX
 PA (INCYTE GENOMICS INC.
 XX
 PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL,
 PI Yu JY, Thaeon O, Yap PE, Ameshey SR, Dam TC, Liu TF, Gerstin EH,
 PI Paralta CH, Lewis SA, Chen A, Marwaha R, Lan RX, Urashka ME,
 PI Kristnam SR, Kolluru V, Panesar IS;
 XX
 XX WPI; 2003-636732/60.
 DR P-PSDB; ADE31156.
 PT
 PT New human diagnostic and therapeutic polynucleotides and polypeptides,
 PT useful for diagnosing, treating or preventing e.g. leukemia, brain
 PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke
 PT or Alzheimer's.
 XX
 PS Claim 1; SEQ ID NO 100; 634bp; English.
 XX
 CC The invention relates to a novel isolated human diagnostic and
 CC therapeutic polynucleotide (designated ditbp). The novel ditbp
 CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
 CC base pairs fully defined in the specification; a polynucleotide
 CC comprising a naturally occurring polynucleotide sequence at least 90%
 CC identical to the ditbp polynucleotide; a polynucleotide complementary to
 CC the ditbp polynucleotide or its polynucleotide which is at least 90%
 CC identical; or an RNA equivalent of any of the polynucleotides mentioned
 CC above. The ditbp polynucleotides have the following activities:
 CC antileukemic, antineoplastic, antitumor, antiproliferative, antidiabetic,
 CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
 CC tranquilizer, osteopathic, anticholinergic, antineuritic, cytoskeletal,
 CC hepatologic, virucide, haemostatic, anti-HIV, antithyroid, thymolitic,
 CC dermatologic, antibacterial, fungicide, antiparasitic, anticonvulsant,
 CC thrombolytic, anticoagulant, anorectic, vasotropic, and anticancer. The
 CC novel DITBP polynucleotides polypeptide can be used in gene therapy and
 CC protein replacement therapy. The ditbp polynucleotides or DITBP
 CC polypeptides are useful for diagnosing, preventing or treating diseases
 CC associated with the expression of human molecules. In particular, these
 CC diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain
 CC cancer), breast cancer, cervix cancer, bone cancer, liver cancer, lung
 CC cancer) or other cell proliferative disorders (e.g. atherosclerosis,
 CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
 CC thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
 CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
 CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
 CC viral, bacterial, fungal or parasitic infection), developmental disorders
 CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
 CC thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic
 CC disorders (e.g. hypercholesterolemia, hypoglycaemia, diabetes,
 CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic
 CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
 CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
 CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),
 CC transport disorders (e.g. akinesia or multidrug resistance), or
 CC connective tissue disorders (e.g. Paget's disease or rickets). This
 CC polynucleotide sequence represents one of the human ditbp DNA sequences
 CC of the invention.
 XX
 SQ Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;

Pred. No.: 2,99e-102 Length: 3358
 Score: 995.50 Matches: 250
 Percent Similarity: 44.82% Conservative: 1
 Best Local Similarity: 44.64% Mismatches: 7
 Query Match: 71.88% Indels: 307
 DB: 10 Gaps: 2
 US-10-649-273-2_COPY_148_414 (1-267) x ADE31345 (1-3358)
 QY 10 LeuThraAsnIyValGluPheProPheLeuValLeuLeuIleSerGlyGlyHsCysLeu 29
 DB 3067 GTGACCAATTAAGTAAATTCCTTTTATGTTCT-TTAATTTCTGAGGTCACTGCTG 3009
 QY 30 LeuAlaLeuValGlyGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAla 49
 DB 3008 TGGGCAATTGTTCAAGAGATTTCAGATTTCTGCTTCTTGAAAGTC-TTGAACATAGCA 2950
 QY 50 ProGlyAspMet----- 53
 DB 2949 CCAAGTGACATCGTTGACAGATAATTAAAGATTTCATCTTTTGTATGT 2890
 QY 53 ----- 53
 DB 2889 TGTCCATTTCACCTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2830
 QY 53 ----- 53
 DB 2829 TTTTATCTTAGTAACTGAAACAAATTCACATATGTGTAGAGAAAAATTAAGAACAGTA 2770
 QY 53 ----- 53
 DB 2769 GTACACGAATTTAATATTTCTTAGCTTTCTTAATAAATGTAAGAGTTCAATATCTGA 2710
 QY 53 ----- 53
 DB 2709 CATTAAGGCTGAATAGTTTGACATACAGTTATGTATTTGCCAAATATATGTATGTAA 2650
 QY 53 ----- 53
 DB 2649 AGAAGCTGCTGTAAACTAATACATACATGCAAAAAAGTAAATTAAGAAATTAATATAGTA 2590
 QY 53 ----- 53
 DB 2589 TTAACATAAGACATTAAGATGCAATGACAGAAATTAATGACAAATTACTTACACCA 2530
 QY 53 ----- 53
 DB 2529 CAGACAGGGTCCCCCGACCCCCCTTTGTTAGAAATACACAGAGGCTACTGCATAT 2470
 QY 53 ----- 53
 DB 2469 ATAGAAATACACAAACAGACAGACAGTCTTCCACAGTGAATTAATAGAAATTAAG 2410
 QY 53 ----- 53
 DB 2409 ACAAGTTCTTATTTATGACGTTCAATTAAGACAGTTATGCACTTCAAGCCATT 2350
 QY 53 ----- 53
 DB 2349 CCAACCATTAAGAAAGCAACATAGACAGGGGAGATGATGGCTTATTTGCGGCTC 2290
 QY 53 ----- 53
 DB 2289 ATCATAGGAACAGGGTGTCTGCTTACCTGAATATACCTATATATTTGCGCAAA 2230
 QY 53 ----- 53
 DB 2229 GTATAGCATTTTATTCATTCAGGGGTTTTTTGTTGTTAGTAATTTCAATTTATTT 2170
 QY 54 -----Leu 54
 DB 2169 CCTTGACATTTTGTGTTGACAGTATTTAATTTATGACGTAAATAATATGTTTCTTT 2110

```

QY 55 AsplysValAlaArgArgLeuSerLeuIleuYshIProGluCysSerThrmSerGly 74
Db 2109 GATAG-GTGGCAAGAAGACTTCTTAATATAACATCCAGAGTGTCCACCAATGAGTGT 2051
QY 75 -GlyValAlaIleGluHleuAlaGlyGlnIysAsnArgPheHleuAspIleuYsh 94
Db 2050 GGGGAAAGCCCAATGAAATTTGGCCAAACAGAAATATATATTTCAATTTGACATCAAC 1991
QY 94 OProLeuHleuHleuAlaYshAsnCysAspPheSerPheThrmGlyLeuGlnHleuValThras 114
Db 1990 TCCCTGCAATCATGCTAAATAATTTGATTTTCTTTACTGACCTTCAACAGTTACTGA 1931
QY 114 pLysIleIleuMetIysGlyGluGlyGlu-----GlyIleGluYsh 124
Db 1930 TAAATATATTAATGAAAGAAAGAAAGAGAGTATATTTCTAATTAATTAAGTTGAACA 1871
QY 125 -----GlyIleGluYsh 129
Db 1870 GATTAATATTCCTGATGATGCTTAAATAATAGCTGCTCATTTCTGACAGCTAATGAGAG 1811
QY 129 IyGlnIleuSerSerAlaAlaAspIleuAlaIleuThraValGlnHleuMetAlaCysH 149
Db 1810 GGCATATCTGTTCTTACAGACGACATTTGCTGCAAGTACAGACACAAATGGCAATGTC 1751
QY 149 IseuValIleuArgThrmHleuAlaIleuPheCysIleuGlnArgAspLeuLeuProG 169
Db 1750 ATCTTGTGAAAGAAACACATGGGCTATTTCTTTGTAGACAGAGACTTTGTAACCTC 1691
QY 169 IAsnAsnAlaValIleuValAlaSerGlyValAlaSerAsnPheYrIleuArgA 189
Db 1690 AAAATAAGCAGTACTGCTGTCATCTGT-GGTGTGCAAGTAACTTCTATATCCACAGAG 1632
QY 189 IAlaGluIleuLeuThrmAlaThrmIleuThrmIleuLeuCysProPheProAlaGluC 209
Db 1631 CTCTGGAATTTTAAAC-AAACGAAACACAGTGCATTTGTTGTCCTCTCCCAACATAT 1573
QY 209 YshTrAspAsnGlyIleuMetIleuAlaTrpAsnGlyIleuYshArgLeuArgAlaGlyLeuG 229
Db 1572 GCACGTGAATTAAGGACTTATGATGATGCAATGGAATGATTAAGAACTACGTCGTGGCTTGG 1513
QY 229 IyIleuHleuHleuAspIleuGluGlyIleuArgYrGluProIysCysProLeuGlyValAspI 249
Db 1512 GCATTTTACATGACATACAGAGCATCCGCTATGAACCAAAATGTCCTCTTGGAGTAGACA 1453
QY 249 IeSerIyGluValGlyGlnAlaSerIleuValProGlnLeuIyMetGluIle 267
Db 1452 TATCAAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGAGATA 1397

```

```

RESULT 12
ABQ75508
ID ABQ75508 standard; DNA; 1572 BP.
AC ABQ75508;
DT 07-NOV-2002 (first entry)
XX
DE Murine siatologycoprotease-like gene sequence SEQ ID NO:7.
XX
KW Murine, mouse; protease; calcium activated neutral protease type 5;
KW CAPN5; trypsinase 4; siatologycoprotease; enzyme; genetic disease;
KW neurological; neuropsychological; psychotic illness; transgenic animal;
KW gene; ds.
OS Mus musculus.
XX
PN M0200245491-A2.
XX
PD 13-JUN-2002.
XX
PF 05-DEC-2001; 2001WO-USO46405.
XX
PR 06-DEC-2000; 2000US-0251803P.
PR 06-DEC-2000; 2000US-0251820P.

```

```

PR 13-DEC-2000; 2000US-0255971P.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Allen KD, Leviten MW;
XX
DR WPI; 2002-657389/70.
XX
PT Novel transgenic animal, comprising a disruption in protease target gene,
PT is useful for identifying agents that ameliorates a phenotype associated
PT with a disruption in a protease target gene.
XX
PS Example 3; Fig 7; 62pp; English.
XX
CC The present invention describes a non-human transgenic animal (I)
CC comprising a disruption in a protease target gene (PG) selected from
CC calcium activated neutral protease type 5 (CAPN5) gene, trypsinase 4 gene
CC and siatologycoprotease-like gene. Also described is a targeting construct
CC (II), comprising a first polynucleotide sequence homologous to at least a
CC first portion of PG, a second polynucleotide sequence homologous to at
CC least a second portion of PG and a selectable marker. (II) is useful for
CC producing a transgenic mouse comprising a disruption in a protease target
CC gene, by introducing (II) into a cell, introducing the cell into a
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC useful for identifying an agent that modulates the expression or function
CC of a protease target gene, by administering an agent to (I) and
CC determining whether the expression or function of the disrupted protease
CC target gene in (I) is modulated. (I) is also useful for testing the
CC efficacy of proposed genetic and pharmacological therapies for human
CC genetic diseases, such as neurological, neuropsychological or psychotic
CC illness. The present sequence represents murine siatologycoprotease-like
CC gene sequence, which is used in an example from the present invention
XX
SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

```

```

Alignment Scores:
Score: 2,676-101 Length: 1572
Percent Similarity: 84.21% Matches: 210
Best Local Similarity: 78.95% Mismatches: 38
Query Match: 70.97% Indels: 6
DB: Gaps: 2
US-10-649-273-2_COPY_148_414 (1-267) x ABQ75508 (1-1572)
QY 1 MetGluAlaHleuAlaLeuThrmIleuArgLeuThrmAsnIyValGluPheProPheLeuVal 20
Db 445 ATGAGAGCTCAGCAGCATGATCTTATAGGCTCAACCAATTAAGTATGATTN-CTTTTATAGTT 503
QY 21 LeuLeuIleSerGlyIyHleuHleuAlaLeuValGlnGlyValIserAspPheLeu 40
Db 504 CTTTGTGATTTCTGGCGGACATGCTGCTGTGAGRTTATGTCAGAGTGTTCCTGATTTCCG 563
QY 41 LeuLeuGlyIySerSerLeuAspIleuAlaProGlyAspMetLeuAspIyValAlaArgArg 60
Db 564 CTCTTGGAGAGTCTTTGGACATGACACAGGCGACATCTTGACAAAGGTGGCAAGAA 623
QY 61 LeuSerLeuIleIyHleuHleuAlaProGluCysSerThrmSerGlyIyValAlaIleGluHleu 80
Db 624 CTTTCTTTATATTAACATCCAGAAATGTTTCAACATGAGTGTGGAGAAACCTATAGAACAG 683
QY 81 LeuAlaIyGlnGlyAsnArgPheHleuAspIleuYshProProLeuHleuHleuAlaIy 100
Db 684 TTGGCCAAAGAGCAAGAAATGATTTCAATTTTACATCAATCCATTAAGCAAGATGCTAAG 743
QY 101 AsnCyAspPheSerPheThrmIyLeuGlnHleuValThrasPlyIleIleuMetIyYsh 120
Db 744 AATTCGATTTTCTTTTCAACGAGACTTCAACATATATTAAGTAAAGTAAATTAACACAAAG 803
QY 121 GluIyGlnGluGlyIleGluYshGlnIleuSerSerAlaIleuAspIleuAlaIleu 140

```

804 GAAAAAGAGAGCATTTGAGAGAGGCGCAATCTGTATCATGAGTCAGACATTCCTGCT 863
QY 141 ThValGlnHisThrMetAlaCysHisIleuValIysAspGlyThrHisArgAlaIleuPhe 160
DB 864 GCGGTACAGCATGACAGCGCTGCCTTCCGAAAGAACATCGCGCTATTCCTGTTT 923
QY 161 CysIleGlnArgAspLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal 180
DB 924 TGCAGAGCAAAAATTTGCTCTCTCCAGCTAACCCAGATTAAGTTGATCTCGAGTGT 983
QY 181 AlSerAsnPheTyrIleArgArgAlaIleuGlnIleuThrAsnAlaThrGlnCysThr 200
DB 984 GCAAGTAATCTGTATCATCCGAAAGCATTTGGAATTTGCGCAATTCGAACGCATGACGG 1043
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB 1044 TTGTTGTGTCCTACT-TCAGACTGTGCACTGACAAATGGCATATGATTCAGATGGA 1102
QY 221 TLeuGlnArgLeuArgAlaGlyLeuGlyIleLeu-HisAsp-TLeuGlnGlyIleArgTyrG 240
DB 1103 ATTGGAAGATTACGTGCTGCTGCTTGGCTTTACCATATATATAGAGACATTCGGTTATT 1162
QY 240 IuProIysCysProLeuGlyVal---AspIleSerTyrGluValGlyIuIaSerIleL 259
DB 1163 AACCCAAATGTCCTCTTGAAGTGAAGGATATATCCGAAAGAGTTGGCAGA---AGCTTGC 1219
QY 259 ysvAlProGlnLeu 263
DB 1220 CCATTAATAAGTTA 1233

RESULT 13

AAS84622
ID AAS84622 standard; cDNA; 2734 BP.

AAS84622;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #20426.

Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

MO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001KO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Dmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PSDB; ABG20435.

New isolated polynucleotide and encoded polypeptides, useful in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biodiversity.

Claim 1; SEQ ID NO 20426; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II)

sequences. (I) is useful as hybridisation probes, polymerase chain

reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

and in recombinant production of (II). The polynucleotides are also used

in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC XX

SQ Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;

Alignment Scores:

Pred. No.:	2,61e-57	Length:	2734
Score:	599.00	Matches:	131
Percent Similarity:	71.57%	Conservative:	10
Best Local Similarity:	66.50%	Mismatches:	22
Query Match:	43.25%	Indels:	34
DB:	5	Gaps:	6

US-10-649-273-2_COPY_148_414 (1-267) x AAS84622 (1-2734)

QY 71 ThrMetSerGlyGlyLeuAlaIleGluHisLeuAlaIleGlnGlyAsnArgPheHispe 90

DB 206 ACCCTGTCAGGGGGGCGCT-----AGATTCCAT----- 235

QY 91 AspIleLysProProLeuHisHisAlaLys-----AsnCys----- 102

DB 236 -----CTATTTCCTCCAGTGAAGATGACGCTGTATGAGAGCGTGCAGCA 283

QY 103 -----AspPheSerPheThrGlyLeuGlnHisValThrAsp 114

DB 284 TCTGCTTCGGAGAGCCCGAGAGACTTTTACTCATGCGGAGGCAAAATGGAGATGAG 343

QY 115 LysIleIleMetLysLeuGluLysGluGlu-----GlyTLeuGluGlnIleLeu 132

DB 344 GCTTTACATGCGCAGAGAGGAGCCAGAGAGAGGGGGGATTTGAGAGAGGGCAATCTCG 403

QY 133 SerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLys 152

DB 404 TCTTCAGGAGCAGACATTCCTCCACAGTACGACACACAAATGTCATCTTGGAAA 463

QY 153 ArgThrHisArgAlaIleLeuPheCysLysGlnAspLeuLeuProGlnAsnAla 172

DB 464 AGAACACATCGGGCTATTCGTTTGTATGAGAGAGACTTGTACTCAAAATATATGCA 523

QY 173 ValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIle 192

DB 524 GTACTGTTGCACTGAGTGTCCAGATTAATCTTATATCGGAGACTCTCGAAATT 583

QY 193 LeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsn 212

DB 584 TTAACAAACGCAACACAGTGCCTTGTGTCTCTCCAGACTATGCACTGTAAT 643

QY 213 GlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHis 232

DB 644 GGCATTATGAATGATGATAATGTATGAAAGATACAGTGCCTGCGGCAATTTACAT 703

QY 233 AspIleGluGlyIleArgTyrGluProLys-----CysProLeuGlyVal 247

DB 704 GACATPAGAGCATTCGCTATGACCAAAATATGTCCTCTTACAGGCTG 754

RESULT 14

ADL86725 standard; DNA; 371 BP.

XX


```
AC ADL86725;
XX
XX 20-MAY-2004 (first entry)
XX
XX DNA up-regulated in murine multipotent progenitor cells Seqid 3118.
XX
XX gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
XX
XX Mus sp.
XX
XX WO2003093445-A2.
XX
XX 13-NOV-2003.
XX
XX 05-MAY-2003; 2003WO-US014114.
XX
XX 03-MAY-2002; 2002US-0377383P.
XX
XX (STOW-) STOWERS INST MEDICAL RES.
XX
XX Li L;
XX
XX WPI; 2004-022656/02.
XX
XX Classifying an unknown multi-lineage affiliated gene comprises isolating
XX expressed nucleic acid sequences from the discrete cell sub-populations.
XX
XX Claim 8; SEQ ID NO 3118; 123bp; English.
XX
XX This invention relates to a novel method for predicting gene potential by
XX associating nucleic acid sequences of unknown function with particular
XX sub-population profiles. Specifically, it refers to classifying an
XX unknown multi-lineage affiliated gene by collecting hybridisation data to
XX develop a gene expression map, in order to determine the discrete sub-
XX population where it is expressed. The present invention describes methods
XX for predicting the lineage commitment of genes associated with the self-
XX renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX referred to as bone marrow stem cells populations. As such, these methods
XX can be used to identify associated multi-lineage affiliated genes and
XX hence the underlying molecular mechanisms in physiological haematopoietic
XX development. This polynucleotide sequence is DNA associated with a murine
XX MPP sub population of cells of the invention.
XX
XX SQ Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1.18e-43 Length: 371
XX Score: 468.00 Matches: 92
XX Percent Similarity: 78.86% Conservative: 5
XX Best Local Similarity: 74.80% Mismatches: 26
XX Query Match: 33.79% Indels: 0
XX DB: Gaps: 0
XX
XX US-10-649-273-2_COPY_148_414 (1-267) x ADL86725 (1-371)
XX
XX 137 AspllelaaathrrValaGlnHsThweraCysHsleuVallyeArGThrHisArg 156
XX 2 GACATTGCTGCTGGCGGTCACACATGCACAGCGTCCACTTGCCAAAGAAACACATGCC 61
XX 157 AlalleuPheCyLyvGlnArgAspleuPProGlnaenAsnAlaValleuVala 176
XX 62 GCATATTCGTGTTTCCAGCAGAAATTTGCTCTCTCCAGCTAACGCGATTAAGTTGTA 121
XX
XX 177 SerGlyGlyValAlaSerAsnPhetyrTleArgArgAlaLeuGluIleuThraAsnAla 196
XX 122 TCTGGAGAGTGTTCGAAAGTAATCTGTACATCCGAAAGCATTTGGAATTTGCCAAATGCA 181
XX 197 ThGlnGysThraLeuLeuCyvProProArgLeuCyvStrraPargGlylleMeclle 216
XX
XX
```

```
DB 182 ACGCAGTCACGTTGTTGTCACCTCCAGACGTGCACTGACATGACATGATGAT 241
XX
XX 217 AATTPaenglyYllegluArgleuArgAlaGlyLeuGlyYlleuHisAsp1leglugly 236
XX 242 GCATGAAATGGAATTAAGTAAGATACGTGCGNGCTTGCTTTTACATGANTATGAAGAC 301
XX
XX 237 lleaGTYrGluProLyvCyvProLeuGlyValaAsp1leSerLyvGluValaGlyAla 256
XX 302 ATCCNNTATGNAACCAAAATNTCNTTTTGAGAGTACATATTCAGAGAAATTCAGAGAGCT 361
XX
XX 257 Ser1leLyv 259
XX 362 GCCATPAAAA 370
XX
XX DB
XX
XX RESULT 15
XX ADL86726
XX ID ADL86726 standard; DNA; 371 BP.
XX
XX AC ADL86726;
XX
XX 20-MAY-2004 (first entry)
XX
XX DNA up-regulated in murine multipotent progenitor cells Seqid 3119.
XX
XX gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
XX
XX Mus sp.
XX
XX WO2003093445-A2.
XX
XX 13-NOV-2003.
XX
XX 05-MAY-2003; 2003WO-US014114.
XX
XX 03-MAY-2002; 2002US-0377383P.
XX
XX (STOW-) STOWERS INST MEDICAL RES.
XX
XX Li L;
XX
XX WPI; 2004-022656/02.
XX
XX Classifying an unknown multi-lineage affiliated gene comprises isolating
XX expressed nucleic acid sequences from the discrete cell sub-populations.
XX
XX Claim 8; SEQ ID NO 3119; 123bp; English.
XX
XX
XX This invention relates to a novel method for predicting gene potential by
XX associating nucleic acid sequences of unknown function with particular
XX sub-population profiles. Specifically, it refers to classifying an
XX unknown multi-lineage affiliated gene by collecting hybridisation data to
XX develop a gene expression map, in order to determine the discrete sub-
XX population where it is expressed. The present invention describes methods
XX for predicting the lineage commitment of genes associated with the self-
XX renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX referred to as bone marrow stem cells populations. As such, these methods
XX can be used to identify associated multi-lineage affiliated genes and
XX hence the underlying molecular mechanisms in physiological haematopoietic
XX development. This polynucleotide sequence is DNA associated with a murine
XX MPP sub population of cells of the invention.
XX
XX SQ Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1.18e-43 Length: 371
XX Score: 468.00 Matches: 92
XX Percent Similarity: 78.86% Conservative: 5
XX Best Local Similarity: 74.80% Mismatches: 26
XX Query Match: 33.79% Indels: 0
XX
```


DB: 12 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-2-267) x ADL86726 (1-3-71)

QY	137	AspIleAlaIaIaThrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArg	156
Db	2	GACATTGCTGCTGGCGGTACAGCATGCACAGCGTGGCACTTGCGAAAGAAACATCGC	61
QY	157	AlaIleLeuPheCysIleGlnArgAspLeuProGlnAsnAsnAlaValLeuValAla	176
Db	62	GCTATTCTGTTTTGGACAGCAGAAATTTTGCTCTCCAGCTAACCCAGATTAGTTGTA	121
QY	177	SerGlyGlyValAlaIaSerAsnPheTyrIleArgArgAlaLeuGluIleuThrAsnAla	196
Db	122	TCTGAGGGGTGTGCAGTAAGTACTGTGCATCCGAAAGCATTTGGAAATTTGCGAAATGCA	181
QY	197	ThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIle	216
Db	182	ACGCAGTGCACCTGTGTGTGTCCACTCCAGACTGTGCACTGCATGCATGGCATCATGATT	241
QY	217	AlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleuHisAspIleGluGly	236
Db	242	GCATGGATATGGATTGAATGAATGATTACCTGCAGCTTGGCCTTTACTGAGNNTAGAAAGAC	301
QY	237	IleArgTyrGluProLysCysProLeuGlyValAspIleSerTyrGluValGlyGluAla	256
Db	302	ATCCNNTATGNAACAAATATCTCNTTTTGAGATGAGATCATTTCCAGAAAGTTGCAGAAGCT	361
QY	257	SerIleLys	259
Db	362	GCCTATPAA	370

Search completed: February 16, 2005, 13:58:48
Job time : 525.461 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:19:37 ; Search time 158.749 Seconds
(without alignments)
2752.056 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHATRLRLNKNVEFPFLV.....DISKVEGASIKVPLQKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 81818359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-Q=/cgn2_1/USPRO.spool/US10649273/runat_14022005_114704_16425/app_query.fasta_1.1429
-DB=Issued Patents NA -QEMT=fastcap -SUFFIX=rml -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NOR=ext -HEAPSIZE=500 -MINDEN=0 -MAXLEN=200000000
-USBR=US10649273 @CGN_1_1_292 @runat_14022005_114704_16425 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	1416	4	US-09-774-528-177 Sequence 177, App
2	1385	100.0	1526	4	US-10-067-443-23 Sequence 23, Appl
3	1385	100.0	2197	4	US-10-067-443-1 Sequence 1, Appl
4	1358.5	98.1	1387	4	US-10-067-443-21 Sequence 21, Appl
5	824.5	59.5	14364	4	US-10-067-443-20 Sequence 38, Appl
6	308.5	22.3	94750	4	US-09-596-002-38 Sequence 806, App
7	307	22.2	1053	4	US-09-540-236-806 Sequence 884, App
8	291.5	21.0	1059	4	US-09-252-991A-884 Sequence 801, Appl
9	291.5	21.0	1206	4	US-09-252-991A-801 Sequence 1, Appl
10	289.5	20.9	1830121	4	US-09-557-884-1 Sequence 1, Appl
11	289.5	20.9	1830121	4	US-09-643-990A-1 Sequence 6612, Ap
12	282	20.4	996	4	US-09-902-540-6612

C	13	282	20.4	2582	4	US-09-902-540-503	Sequence 503, App
	14	280.5	20.3	1074	4	US-09-543-681A-2341	Sequence 2341, Ap
	15	270.5	19.5	1315	1	US-08-087-797-1	Sequence 1, Appl
	16	268.5	19.4	1032	4	US-09-489-039A-2050	Sequence 2050, Ap
	17	261	18.8	1092	4	US-08-987-121A-5	Sequence 2955, Ap
	18	254	18.3	1008	3	US-08-987-121A-5	Sequence 5, Appl
	19	254	18.3	1011	3	US-08-987-121A-5	Sequence 3, Appl
	20	250	18.1	1006	3	US-08-961-083-51	Sequence 51, Appl
	21	250	18.1	1006	4	US-09-536-784-51	Sequence 51, Appl
	22	250	18.1	10974	3	US-08-961-527-214	Sequence 214, App
	23	248	17.9	1011	3	US-09-066-512-1	Sequence 1, Appl
	24	246	17.8	1011	4	US-09-583-110-2196	Sequence 2196, Ap
	25	241	17.4	1011	4	US-09-107-433-1618	Sequence 1618, Ap
	26	236.5	17.1	640681	4	US-09-790-988-1	Sequence 727, App
	27	236	17.0	822	4	US-09-10-279-727	Sequence 1551, Ap
	28	236	17.0	1101	4	US-09-134-000C-1551	Sequence 3985, Ap
	29	236	17.0	3993	3	US-09-110-279-3985	Sequence 1072, Ap
	30	233	16.8	1107	3	US-09-134-001C-1072	Sequence 1, Appl
	31	220	15.9	1026	3	US-09-149-624-1	Sequence 794, App
	32	218	15.7	3064	3	US-09-221-017B-794	Sequence 1, Appl
	33	218	15.7	1230025	4	US-09-198-452A-1	Sequence 1, Appl
	34	218	15.7	1230230	4	US-09-438-185A-1	Sequence 102, App
	35	216	15.6	15249	4	US-08-956-171E-102	Sequence 102, App
	36	216	15.6	15249	4	US-08-781-986A-102	Sequence 2, Appl
	37	215.5	15.6	4403765	3	US-09-103-840A-2	Sequence 147, App
	38	215.5	15.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
	39	211.5	15.3	11555	4	US-09-602-777A-147	Sequence 1, Appl
	40	207.5	15.0	1664976	4	US-08-916-421B-1	Sequence 1, Appl
	41	207.5	15.0	1664976	4	US-09-692-570-1	Sequence 6, Appl
	42	192.5	13.9	1663	4	US-09-620-312D-6	Sequence 131, App
	43	185.5	13.4	42325	4	US-08-311-731A-131	Sequence 261, App
	44	181.5	13.1	432	4	US-09-328-352-261	Sequence 19, Appl
	45	169	12.2	876	4	US-09-724-623-19	

ALIGNMENTS

RESULT 1

US-09-774-528-177
; Sequence 177, Application US/09774528

; Patent No. 6743619

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Zhou, Ping

; APPLICANT: Goodrich, Ryle

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Ren, Feiyan

; APPLICANT: Zhang, Jie

; APPLICANT: Zhao, Qing A.

; APPLICANT: Yang, Yonghong

; APPLICANT: Xue, Aidong J.

; APPLICANT: Wehrman, Tom

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Wang, Dunrui

; APPLICANT: Dmanac, Radoje T.

; TITLE OF INVENTION: No. 6743619e1 Nucleic Acids and

; FILE REFERENCE: Polypeptides

; CURRENT PILING DATE: 2001-01-30

; NUMBER OF SEQ ID NOS: 441

; SOFTWARE: pc_fl_genes Version 2.0

; SEQ ID NO 177

; LENGTH: 1416

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: CD5

; LOCATION: (205) ..(1305)

US-09-774-528-177

Alignment Scores:

Pred. No.: 2,51e-177 Length: 1416
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-09-774-528-177 (1-1416)

```
QY 1 MetGUAIAHIAIAleuThrlleaRglleuThraPheValIGluPheProPheleuVal 20
DB 502 ATGAGGCTCAGACCTTAATAGTTAGTACCAATTAAGTAATTCCTTTTATGAT 561
QY 21 LeuLeuIleSerGlyGlyYHISCySleuLeuAlaleuValIGInGlyValSerAspPheleu 40
DB 562 CTTTGATTTTCGAGAGTCACTGTCTGTGGCATTTAGTTCAAGAGCTTTCAGATTTTCG 621
QY 41 LeuLeuGlyLySerSerleuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60
DB 622 CTTCTTGAAAGCTCTTTGGACATAGACCAAGTGACATCTTGACAAAGGTGGCAAGAGA 681
QY 61 LeuSerleuIleLyHisProGlyCySerThrMetSerGlyGlyLyAlaIleGluHis 80
DB 682 CTTCTTTAATTAACATCCAGAGTGTCCACCATGAGTGAGGAAAGCCATAGAACAT 741
QY 81 LeuAlaLySGInGlyAsnArgPheHisPheAspIleLyProProleuHisIleAlaLyS 100
DB 742 TTGGCCAAACAAGAAATGATTTCAATTTGACATCAACCTCCCTTGACATCATGTCTAA 801
QY 101 AsnCyAspPheSerPheThrGlyLeuGInHisValThrAspLySleIleMetLySlys 120
DB 802 AATTGTGATTTTCTTTTACTGGACTTGCAACGTTTACATGAATAATATATGAAAG 861
QY 121 GluLySGInGlyGlyIleGluLySGInIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 862 GAAAAAGAGGAAGTATGAGAAAGGGGCAAACTCTGCTTCAGACGACAACTTGCTGCC 921
QY 141 ThrValIGInHisThrMetAlaCySHISleuValLyArgThrHisArgAlaIleleuPhe 160
DB 922 AAGAGACAGACACAATGAGCATGTCAATTTGAAAAGAACATCGGGCTATTCGTTT 981
QY 161 CysLySGInArgAspLeuLeuProGInAsnAlaValIleuValAlaSerGlyGlyVal 180
DB 982 TGTAAAGCAGAGAGACTTGTACTCAAAATATGACATGCTGGTGCATGTGGTGTGC 1041
QY 181 AlaSerAsnPheTyrlleaRgAlaLeuGluIleleuThrAsnAlaThrGInCySThr 200
DB 1042 GCAAGTAACCTTATATCCGACAGAGCTCGAAATTTTAACAAACGCAACACAGTCACT 1101
QY 201 LeuLeuCyProProProProArgleuCySThrAspAsnGlyIleMetIleAlaITTPAsnGly 220
DB 1102 TTGTTGTGTCTCTCCCTCCAGACTATGACATGATTAATGACATTAATGCAATGGAATGAT 1161
QY 221 IlegluArgLeuArgAlaGlyLeuGlyIleleuHisAspIleGlyGlyIleArgTyrglu 240
DB 1162 ATTGAAGACATACGCTGCTGGCTGGCATTTTACATGACATGAAAGGCAATCCGCTATGAA 1221
QY 241 ProLySGyProleuGlyValAspIleSerLySGInValIGlyGluAlaSerIleLyVal 260
DB 1222 CCAAAATGTCTCTTGAGATAGACATATCAAAAAGAGTTGAGAGAGCTTCCATAAAAAGTA 1281
QY 261 ProGInLeuLyMetGluIle 267
DB 1282 CCACAAATTAATAATGAGATA 1302
```

RESULT 2
US-10-067-443-23

; Sequence 23, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1

FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 2,85e-177 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-23 (1-1526)

```
QY 1 MetGUAIAHIAIAleuThrlleaRglleuThraPheValIGluPheProPheleuVal 20
DB 1 ATGAGGCTCAGACCTTAATAGTTAGTACCAATTAAGTAATTCCTTTTATGAT 60
QY 21 LeuLeuIleSerGlyGlyYHISCySleuLeuAlaleuValIGInGlyValSerAspPheleu 40
DB 61 CTTTGATTTTCGAGAGTCACTGTCTGTGGCATTTAGTTCAAGAGCTTTCAGATTTTCG 120
QY 41 LeuLeuGlyLySerSerleuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60
DB 121 CTTCTTGAAAGCTCTTTGGACATAGACCAAGTGACATCTTGACAAAGGTGGCAAGAGA 180
QY 141 ThrValIGInHisThrMetAlaCySHISleuValLyArgThrHisArgAlaIleleuPhe 80
DB 181 CTTTCTTAATTAACATCCAGAGTGTCCACCATGAGTGAGGAAAGCCATAGAACAT 240
QY 161 LeuSerleuIleLyHisProGlyCySerThrMetSerGlyGlyLyAlaIleGluHis 80
DB 181 CTTTCTTAATTAACATCCAGAGTGTCCACCATGAGTGAGGAAAGCCATAGAACAT 240
QY 81 LeuAlaLySGInGlyAsnArgPheHisPheAspIleLyProProleuHisIleAlaLyS 100
DB 241 TTGGCCAAACAAGAAATGATTTCAATTTTGAATCAACCTCCCTTGACATCATGTCTAA 300
QY 101 AsnCyAspPheSerPheThrGlyLeuGInHisValThrAspLySleIleMetLySlys 120
DB 301 AATTGTGATTTTCTTTTACTGGACTTGCAACGTTTACATGAATAATATGAAAAAG 360
QY 121 GluLySGInGlyGlyIleGluLySGInIleleuSerSerAlaAlaAspIleAlaAla 140
DB 361 GAAAAAGAGGAAGTATGAGAAAGGGGCAAACTCTGCTTCAGACGACAACTTGCTGCC 420
QY 141 ThrValIGInHisThrMetAlaCySHISleuValLyArgThrHisArgAlaIleleuPhe 160
DB 421 AAGAGACAGACACAATGAGCATGTCAATTTGAAAAGAACATCGGGCTATTCGTTT 480
QY 161 CysLySGInArgAspLeuLeuProGInAsnAlaValIleuValAlaSerGlyGlyVal 180
DB 481 TGTAAAGCAGAGAGACTTGTACTCAAAATATGACATGCTGGTGCATGTGGTGTGC 540
QY 181 AlaSerAsnPheTyrlleaRgAlaLeuGluIleleuThrAsnAlaThrGInCySThr 200
DB 541 GCAAGTAACCTTATATCCGACAGAGCTCGAAATTTTAACAAACGCAACACAGTCACT 600
QY 201 LeuLeuCyProProProProArgleuCySThrAspAsnGlyIleMetIleAlaITTPAsnGly 220
DB 601 TTGTTGTGTCTCTCCCTCCAGACTATGACATGATTAATGACATTAATGCAATGGAATGAT 660
QY 221 IlegluArgLeuArgAlaGlyLeuGlyIleleuHisAspIleGlyGlyIleArgTyrglu 240
DB 661 ATTGAAGACATACGCTGCTGGCTGGCATTTTACATGACATGAAAGGCAATCCGCTATGAA 720
```

```

OY 241 ProLysCyseProLeuGlyValAspIleSerLysGlyValGlyGlnAlaSerIleLysVal 260
Db 721 CCAAAATGTCCTTTGGAGTAGACATATCAAAAGAGTTGGAGAGCTTCATAAAAAGTA 780
OY 261 ProGlnLeuLysMetGluIle 267
Db 781 CCACATTAATAAATGGAGATTA 801

RESULT 3
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-067-443-1

Alignment Scores:
Pred. No.: 5 23e-177 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-1 (1-2197)
OY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db 672 ATGGAGAGCTCATGACCTTACTATTAGGTGACCAATTAAGTAGAATTCCTTTTACTGTT 731
OY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 732 CTTTGGATTTCTCGAGGCTCACTGCTCTGTGGCATTAAGTTCAAGAGATTTCAGATTTCTG 791
OY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 792 CTTCTGAAAGCTTTTGACATAGCACAGGTGACATGCTTGACAGAGTGGCAAGAGA 851
OY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysValIleGlnHis 80
Db 853 CTTTCTTATATAAACAATCCAGAGTGTCCACACATGAGTGGTGGGAAACCATAGAACAT 911
OY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 912 TTGGCCAAACAAAGAAATAGATTTCATTTTACATCAAAACCTCCCTGCATCATGCTAA 971
OY 101 AsnGlyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 972 AATTGATTTTCTTTTACTGCACTTCACACCTTACTGATAAATAATAATGAAAAG 1031
OY 121 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db 1032 GAAAAAGAGAGAGCTATTGAGAAAGGGCAAACTCTGCTTCACGACAGACATCTGCTCC 1091
OY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160

```

Db	Accession	Gene	Length	Score	Best Local Similarity	Query Match	DB
Db	1092	ACNGRACAGACAAATGGCATCTCTGTGTAAAGAACACATCGGGCTATTCTGTTT	1151				
Qy	161	CyLysglnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal	180				
Db	1152	TGTAAAGCAGAGACTGTGTACTCAAAATTAATGCACTAGTGGTGCATGGTGGTGC	1211				
Qy	181	AlaSerAsnPhenylTyrIleArgGlnAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr	200				
Db	1212	GGAAATGATCTTCTATATCCGACAGAGCTCTGGAAATTTTAACTAAACGACACAGTGCCT	1271				
Qy	201	LeuLeuGlyProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly	220				
Db	1272	TTGTTGTGCTCTCTCCCGACATGATGACTGATTAATGCACTTATGATTCATGGAAATGCT	1331				
Qy	221	IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu	240				
Db	1332	ATTGAAAGACTAGTGTGCTGGCTGGCATTTTACATGATGACATAGAAAGCATCCGCTATGAA	1391				
Qy	241	ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal	260				
Db	1392	CCAAATATGCTCTCTGGAGTACATATCAATCAAAAGATTGGAGAGCTTCATTAAGTA	1451				
Qy	261	ProGlnLeuLysMetGlnIle 267					
Db	1452	CCACATTTAAATGAGATA 1472					
RESULT 4							
US-10-067-443-21							
; Sequence 21, Application US/10067443							
; Patent No. 6642041							
; GENERAL INFORMATION:							
; APPLICANT: Bristol-Myers Squibb Company							
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN							
; TITLE OF INVENTION: SPINAL CORD, MP-1							
; FILE REFERENCE: D0073 NP							
; CURRENT APPLICATION NUMBER: US/10/067,443							
; CURRENT FILING DATE: 2002-02-05							
; PRIOR APPLICATION NUMBER: US 60/266,518							
; PRIOR FILING DATE: 2001-02-05							
; PRIOR APPLICATION NUMBER: US 60/282,814							
; PRIOR FILING DATE: 2001-04-10							
; NUMBER OF SEQ ID NOS: 71							
; SOFTWARE: PatentIn version 3.0							
; SEQ ID NO 21							
; LENGTH: 1387							
; TYPE: DNA							
; ORGANISM: homo sapiens							
US-10-067-443-21							
Alignment Scores:							
Pred. No.: 9.58e-174 Length: 1387							
Score: 1358.50 Matches: 266							
Percent Similarity: 91.10% Conservative: 0							
Best Local Similarity: 91.10% Mismatches: 1							
Query Match: 98.09% Indels: 25							
DB: 4 Gaps: 1							
US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-21 (1-1387)							
Qy	1	MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal	20				
Db	465	ATGAGAGGCTCAGCACTTACTATTAGTTGACCAATAAAGTAGAATTTCTTTTATGTT	524				
Qy	21	LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValIleSerAspPheLeu	40				

|||||
Db 645 CTTCTTAAATAAACAATCCAGAGTGTCCACCAATGAGTGGGAAAGCCATAGAGCAT 704
Qy 81 LeuAlaIyGsiGngIyAserAphEhiSphEapIleIySProProLeuHiShiSaIalyS 100
Db 705 TTGGCCAAACAAGAAATAGATTTCATTTTGACATCAACCTCCCTTGATCATGCTAAA 764
Qy 101 AasnCyAapPheSerPheThGlyLeuGlnHisValThrAplysIleIleMetIySlyS 120
Db 765 AATTGTGATTTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATAATATGAAAAAG 824
Qy 121 GluIyGsiGngIyIle----- 126
Db 825 GAAAAAGAGAGAGATATTTCTAATTAGTAAAGTTGAACAGATAAATATCTCGATTTG 884
Qy 127 -----GluIyGsiGngIleIleuSerSeraIa 135
Db 885 TGCCATAAAATAGCTCTCATTTCTGACAGTATGAGAAAGGGCAATCCTGCTTCAGCA 944
Qy 136 AlaAapIleAlaIaThrValGlnHisThrMetAlaCysHisIleuValIySArgThrHis 155
Db 945 GCAAGCATGCTGCGACAGTACAGACCAATGSCATGTCATCTGTGAAAAAGAACACAT 1004
Qy 156 ArgAlaIleIleuPheCysIyGlnArgAapIleuProGlnAasnAapAlaValIeuVal 175
Db 1005 CGGCTATTCTGTTTGTAGACAGAGACATTGTACCTCAAAATATATGACATGCTGTT 1064
Qy 176 AlaSerGlyIyValAlaSerAapPheTyIleArgArgAlaIeuGlnIleuThrAen 195
Db 1065 GCATCTGGTGGTGGCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTAAACAAAC 1124
Qy 196 AlaThrGlnCysThrIleuIleuCysProProAagIeuCysThrAapAenIyIleMet 215
Db 1125 GCAACACAGTGCATCTTGTGTGCTCTCCCAACATATGACATATATGCACTTAAG 1184
Qy 216 IleAlaTPaengIyIleGlnArgIeuArgAlaGlyIleuGlnIleuHiSaapIleGln 235
Db 1185 ATTGCATGGAATGATGATGAAAGACTACGTGCTGGCTTTGACATGACATAGAA 1244
Qy 236 GlyIleArgTyIyGlnProIySAsPProIeuGlyValAapIleSerIySglnValGlyGln 255
Db 1245 GGCACTCCGCTATGAACCAAAATGTCTCTTGAGATGACATATCAAAAGAGTTGAGAA 1304
Qy 256 AlaSerIleIyValProGlnIleuIySmetGluIle 267
Db 1305 GCTTCCATAAAGTACCAATTAATAATGAGATA 1340
RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, ME-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20
Alignment Scores:
Pred. No.: 1,34e-99 Length: 14364
Score: 824.50 Matches: 240

Percent Similarity: 26.73% Conservative: 0
Best Local Similarity: 26.73% Mismatches: 2
Query Match: 59.53% Indels: 657
DB: 4 Gaps: 4
US-10-649-273-2_copy_148_414 (1-267) x US-10-067-443-20 (1-14364)
Qy 1 MetGlnAlaHisAlaIleuThrIleArgIleuThrAasnIyValGlnPheProPheIeuVal 20
Db 10845 ATGAGAGCTCAATGCACTTATCTATAGTTGACCAATAGATGAAATTTCTTTTATGTT 10904
Qy 21 LeuIleIleSerGlyIyHisCysIleuIeuAlaIeuValGlnGlyValSerAapPheIeu 40
Db 10905 CTTTGAATTTCTGAGAGTCACTGTCTGTGGCATTAGTTCAAGAGTTTCAAGATTTCTG 10964
Qy 41 LeuIeuGlyIySerIeuAapIleAlaProGlyAapMetIeuAplysVal----- 57
Db 10965 CTTCAATGGAAGCTTTTGACATAGACCAAGGTGACATGCTTGAACAAAGT-AAATTAAGA 11023
Qy 57 ----- 57
Db 11024 TTAATTTCTCAATCTTTTGTATGTTGTCATTTCATTAAGTAACTGAATAATGATGTC 11083
Qy 57 ----- 57
Db 11084 TACACCATTCACCTAAATTTTCTGAATTTTATCTAGTAAACTGAAAAATTCACAT 11143
Qy 57 ----- 57
Db 11144 AAGTGAGAAAAAATAGAAAGATGATACACAAATTTATATCTTACGCTTTCTTAATA 11203
Qy 57 ----- 57
Db 11204 AAATGTAAGAGTTCATATCTGATCAATAAGCTGAATAGTTGCAGATACATTAAG 11263
Qy 57 ----- 57
Db 11264 TATTTGGCCAAATATGATATGTAAGAAAGCGTCTGTAACATACTAGCGCAAAAAA 11323
Qy 57 ----- 57
Db 11324 GGTAATAATAGAAATATATATATAGATTAACATTAAGACATTAAGATGCAATGACAGAA 11383
Qy 57 ----- 57
Db 11384 TTAATCACAATAATTAATTACACACAGACAGAGTCCCCCAGACCCCTTTGTTTAG 11443
Qy 57 ----- 57
Db 11444 AATACTACAGAGGCTACGCTATATATAGAAAAACAAACAAACAAACAAACAAACATG 11503
Qy 57 ----- 57
Db 11504 CTTCCACAGTGAATATATAGGAATGATAGCAAGTCTTATATATGAGCTTCATCAT 11563
Qy 57 ----- 57
Db 11564 AAGCAGTATTTGCACTTCAAGCCCATTTTCCAAACAAATAGAAAGCAAAACATPAGACAG 11623
Qy 57 ----- 57
Db 11624 GGGCAGTATTTGGCTTTATTTGTTGGGTGATCATTAAGAAACAGGTTGTCTGTAC 11683
Qy 57 ----- 57
Db 11684 TGAATATCAGTATAGTATATATTTGCCAAAGTATAGCATTTTATTCATTCAGGGGTT 11743
Qy 57 ----- 57
Db 11744 TTTTGTGTTAGTAAATTTCAATTTATTCCTTGACATCTTTTGCTTTCACAGATTTT 11803
Qy 58 -----AlaArgArgIeuSerIleIle 64
|||||

```

Db 11804 AATTATAGACTCTAAATAATATGTTTCTTGATAGTGGCAAGACCTTTCTTAATA 11863
QY 65 LysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHisIleuAlaIleuGln 84
Db 11864 AATCATCCAGAGTGTCTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACGA 11923
QY 85 GlyAsnArgPheHisPheAspIleLeuSerProLeuHisHisAlaIleuAsnCysAspPhe 104
Db 11924 GGAATATGATTTCACTTTGACATCAAACTCCCTTGACATGCTGTAATAATGTGATTTT 11983
QY 105 SerPheThrGlyLeuGlnHisValIleuThrAspIleIleMetLeuGlyGlyLeuGlu 124
Db 11984 TCTTTTACTGACTTCAACACCTTACTGATTAATAATATGATAAAGGAAAAAGGAGAA 12043
QY 124 ----- 124
Db 12044 GGTATATTTCTAATATGTAAGTGAACAGATAAATATTCCTGATGTGCTTAAAAATA 12103
QY 125 -----GlyIleGlyLeuGlyGlnIleLeuSerSerAlaAlaAspIleAl 139
Db 12104 GCTGCTCATTTCTGACAGGATTTGAGAAAGGCGAAATCCGTCTTCAAGCAGCAGATTC 12163
QY 139 AlaIleThrValGlnHisIleuMetAlaCysHisIleuValIleuAspThrHisArgAlaIle 159
Db 12164 TGCCACAGTACAGCACACAAATGGCATGTCATCTTGAAAGAAAGAACACATCGGGCTATTCT 12223
QY 159 uPheCysLeuGlnHisArgAspLeuLeuProGlnAsnAsnAlaValIleu----- 174
Db 12224 GTTTTGTAGCAGAGAGACTGTGTACTCTCAAAATATGACATGCTGTAAGTTTATCTTC 12283
QY 174 ----- 174
Db 12284 ATTTTATATGTAATAGTACACTTGGCAATANGTACTTTTCCACAGACTTGACCTTG 12343
QY 174 ----- 174
Db 12344 TGTTAGAGTGAACGATCTTATGCTTATGCTTATGCTAGCCCTGACAGTATGAATATG 12403
QY 174 ----- 174
Db 12404 GATGAGAAAGACTAACAGCCATTTCTGTGATGATTTGGTAGCTTATGAGCAGCTGTA 12463
QY 174 ----- 174
Db 12464 TAGCTTCTATGACATATGCTAATTTTGCATCTTCTGTGATTTAAAGAGGCTT 12523
QY 174 ----- 174
Db 12524 ACATTAAGAAAGTAATATGACATGCTATGACTATTTTGAAGAAATAGGTGATTT 12583
QY 174 ----- 174
Db 12584 CCTTCATCCTTGAATGAAATCCCTTGTGTTGTTTGTATTAATAGCCAGTCAATTT 12643
QY 174 ----- 174
Db 12644 AGCAGTGGAGGTGTATTCCAACTTTCGTGACACTAAATGTTGATTAAGTTCTGATATC 12703
QY 174 ----- 174
Db 12704 CACTATATTTGACAGCCAAATCCCTTAATGTGCTTAAAGCCTTGACAAATCTCT 12763
QY 174 ----- 174
Db 12764 GTTTAAGTATCTTAATCTTATTCATTTTAAATATTAATTAAGTGAAGAAATGTT 12823
QY 174 ----- 174
Db 12824 TAAATGTAATATCATATGATGAATTTTACATGATATCAAGAAATATTTTTCAGA 12883
QY 174 ----- 174
Db 12884 GTTATGTAATAATGACAAATAATATAAATTTACAGGCTCTAAATATGCTACTATGTA 12943

```

```

QY 174 ----- 174
Db 12944 TTGAATATTAATTAATAATATTTAGATGAAGGTGGAGAAATATACAAAAATGCT 13003
QY 174 ----- 174
Db 13004 AGTAATGTTGATAGCTATTAAGAAATTAATAGTAATTTTCTTCCAAATTTTATAC 13063
QY 174 ----- 174
Db 13064 ATAGATATGTCATCTGCCATTACCATCTCAAAATGGATAGTTATATGTTTAATG 13123
QY 174 ----- 174
Db 13124 CTGATATTTTCTCCAGTTTAAATTAGCAGCTTGTTTCATATCCATATATGATGTTAT 13183
QY 175 -----ValAlaSerGlyValAlaSerAsnPheTyrIleAs 187
Db 13184 TTGTTTCTCAATTCCTTCAGCTTGCATCTGTGTGTGTCGCAATTAATCTTATATCCG 13243
QY 187 gAlaAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAla 207
Db 13244 CAGAGCTCGAAATTTTAACAAACCAACACAGTGCACCTTGTGTGCTCTCCAG 13303
QY 207 gLeuCysThrAspAsnGlyIleMetIleAlaTrp----- 218
Db 13304 ACTATGCACTGATTAATGGAATTAATGATTCATGTAAAGCAGAGATATACGTGCTTAC 13363
QY 218 ----- 218
Db 13364 TCATACTATGTAATATTAATTCGCATTTTATCATACTAAGCTTCTTCAGATCT 13423
QY 219 -----AsnGlyIleGluArgLeuArg 226
Db 13424 TGAGCTATGATTTTATTTATGCTTCTTATTAAGATGATGATTAAGAGCTAGTG 13483
QY 226 lAglyLeuGlyIleLeuHisAspIleGlyGlyIleArgTyrGluProLys 242
Db 13484 CTGGCTTGCGCATTTTATCATGACATGAAGAGCATCGCTATGAACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, B.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAYELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:
; US-09-596-002-38

Alignment Scores:
Pred. No.: 3,21e-20 Length: 94750
Score: 308.50 Matches: 87
Percent Similarity: 49.61% Conservative: 40
Best Local Similarity: 33.98% Mismatches: 101
Query Match: 22.27% Indels: 29

```

```

DB: 4 Gaps: 8
US-10-649-273-2_COPY_148_414 (1-267) x US-09-596-002-38 (1-94750)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThr-----AsnLysValaGluPheProPhe 18
DB 24215 ATGAGAGGACATCTGTAGCACCCTATTGGCCAGTATGACCCCAAGCTTTCCGTTT 24274
QY 19 LeuValLeuLeuIleSerGlyGlyHisCysValLeuAlaLeuValaGlnGlyValaSerAsp 38
DB 24275 GTGTCCCTGTGGTGGGCGGTATACCATGCTGGTCCGATGGTGTGGCGCTG 24334
QY 39 PheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysVala 58
DB 24335 TATCAGATATTGGCGAGTCTATCATATGATGCGGTGGTAATGCTTGAATAAACGGCA 24394
QY 59 ArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIle 78
DB 24395 AAAATGCTCAAACTG---CCCTATCT-----GGTGGCCCAATATC 24433
QY 79 GluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHis 98
DB 24434 GAAAAATTAGCCAAAACGGAACCCACACGCTATGAGCTGCCAAAGCCATGACGAT 24493
QY 99 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValaThrAspLysIleIleMet 118
DB 24494 ---AAAGGCTGATTTTTCCTTTCAGTGCATGGAATAACCCCATTCATATATCAAA 24550
QY 119 LysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIle 138
DB 24551 GACACACCAAGCCCAAGGACCC-----GCCACAGACGACAGCATC 24595
QY 139 AlaAlaThrValaGlnHisThrMetAlaCysHisLeuValaLysArgThrHisArgAlaIle 158
DB 24596 GCCCAAGCTTTGATAGTACCGGTGGTGAATCTTGTGTCAAAATAAGCACCAAGCACATA 24655
QY 159 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaValaLeuValaAlaSerGly 178
DB 24656 CAGATGACAGGCAATTCGCCAG-----CTGGTGTCCGACAGG 24691
QY 179 GlyValaLysAsnAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 198
DB 24692 GGGCTCTCTGCCATCATATGCTACCGCCACCTGACCGAAGCCTCCGCCAATTCAT 24751
QY 199 CysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218
DB 24752 GCGTCGGTACTATGCCCCGACCGAGCTATGACGATATATGTCGATGATCCCTAT 24811
QY 219 AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArg 238
DB 24812 GCTGGCTTTTGTCCGCTCAGCTGTGAACATGCG-----GATGACTTGGC-GGTTCCGCTG 24864
QY 239 TyrGluProLysCysProLeuGlyValaAspIleSerLysGluValaGly 254
DB 24865 TATTC-----CCGATGGGATATGACAGCGCTTGGCCGATTCGCG 24903

```

```

RESULT 7
US-09-540-236-806
; Sequence 806, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 806
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: M.cattarrhalis
; US-09-540-236-806

```

```

Alignment Scores:
Pred. No.: 2,876-31 Length: 1053
Score: 307.00 Matches: 84
Percent Similarity: 48.02% Conservative: 37
Best Local Similarity: 33.33% Mismatches: 91
Query Match: 22.17% Indels: 40
DB: 4 Gaps: 7
US-10-649-273-2_COPY_148_414 (1-267) x US-09-540-236-806 (1-1053)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThr-----AsnLysValaGluPheProPhe 18
DB 355 ATGAGAGGACATCTGTAGCACCCTATTGGCCAGTATGACCCCAAGCTTTCCGTTT 414
QY 19 LeuValLeuLeuIleSerGlyGlyHisCysValLeuAlaLeuValaGlnGlyValaSerAsp 38
DB 415 GTGTCCCTGTGGTGGGCGGTATACCATGCTGGTCCGATGGTGTGGCGCTG 474
QY 39 PheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysVala 58
DB 475 TATCAGATATTGGCGAGTCTATCATATGATGCGGTGGTAATGCTTGAATAAACGGCA 534
QY 59 ArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIle 78
DB 535 AAAATGCTCAAACTG---CCCTATCT-----GGTGGCCCAATATC 573
QY 79 GluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHis 98
DB 574 GAAAAATTAGCCAAAACGGAACCCACACGCTATGAGCTGCCAAAGCCATGACGAT 633
QY 99 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValaThrAspLysIleIleMet 118
DB 634 ---AAAGGCTGATTTTTCCTTTCAGTGCATGGAATAACCCCATTCATATATCAAA 690
QY 119 LysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIle 138
DB 691 GACACGCCCAAGCCCAAGGACCCCT-----GCCACAGACGACAGCATC 735
QY 139 AlaAlaThrValaGlnHisThrMetAlaCysHisLeuValaLysArgThrHisArgAlaIle 158
DB 736 GCCCAAGCTTTGATAGTACCGGTGGTGAATCTTGTGTCAAAATAAGCACCAAGCACATA 795
QY 159 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaValaLeuValaAlaSerGly 178
DB 796 CAGATGACAGGCAATTCGCCAG-----CTGGTGTCCGACAGG 831
QY 179 GlyValaLysAsnAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 198
DB 832 GGGCTCTCTGCCATCATATGCTACCGCCACCTGACCGAAGCCTCCGCCAATTCAT 891
QY 199 CysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218
DB 892 GCGTCGGTACTATGCCCCGACCGAGCTATGACGATATATGTCGATGATCCCTATC 227
QY 219 AsnGlyIleGluArgLeuArgAlaGly-----
DB 952 GCTGGCTTTTGTCCGCTCAGCTGTGAACATGCG-----GATGACTTGGC-GGTTCCGCTG 1011
QY 228 -----LeuGlyIleLeuHisAsp 233
DB 1012 CGATGGGATATGACAGTCTTGTGATTCGATATGAT 1047

```

```

RESULT 8
US-09-252-991A-884/c
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

```



```
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 884
/ LENGTH: 1059
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

Alignment Scores:
Pred. No.: 3,67e-29 Length: 1059
Score: 291.50 Matches: 87
Percent Similarity: 50.41% Conservative: 37
Best Local Similarity: 35.37% Mismatches: 99
Query Match: 21.05% Indels: 23
DB: Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-884 (1-1059)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
DB 711 ATGGAAGGGGCACTGCTGGCCGCGATGCTGGAAGAGACAGCACCGCGGTTCCCGTTCCGTC 652
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39
DB 651 GCCCTGCTGCTTCCGGCGGTGACACCCAGTTGTGCGGGTGAGACGGATATCGGCGCGTAC 592
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
DB 591 CAGTTGCTTGGCGAATCGGTGAGATGCCCGCGGAAGCCTTGACAAAGACCGCCAG 532
QY 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79
DB 531 CTGATCGGCGCTG---GGCTATCCC-----GGTGGTCCGGAATCCGC 493
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAla 99
DB 492 CGCTGCGGAGCGCGGCACTCTGCGCTTCCGTTCCCGCGCGCATGACCGCATCGC 433
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
DB 432 CCGGCGCTGAGCTTCAAGCTTCAAGGAGCTTACCTGACAC---ACTGGCAG 376
QY 120 LysGluLysGlnGluGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
DB 375 CGTTGCGTCAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 325
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
DB 324 CTGGCGTTCCAGACCGCGGTGTGAGACCTGCTGATCAAGTGCCTGCGCGCTTG--- 268
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGly 179
DB 267 -----AAGAGACCGGCTG-----AAGAAC-----CTGGTATGCGCGCGCT 229
QY 180 ValAlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
DB 228 GTCAAGCCCAACACAGGCGCTGCGAGCGGCTGAAAAGATGCTCGGGAATGAAGGGG 169
QY 200 ThrLeuLeuCysProProIleArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsn 219
DB 168 CAGGTGTTTCAAGCGCGCGCTTCTGACCGACATGCGGATATGCGCTACGCC 109
QY 220 GlyIleGluArgLeuAlaGlyIleGluGlyIleLeuHisAspIleGluGlyIleArgTyr 239
DB 108 GGTGCGCAGGCGCTGCTGCGCGC-----CAGCATGACGCGCGCGCATGACCGCTC 58
QY 240 GluProLysCysProLeu 245
DB 57 CAGCGCGCTGCGCGCATG 40
```

```
RESULT 9
US-09-252-991A-801
/ Sequence 801, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 801
/ LENGTH: 1206
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-801

Alignment Scores:
Pred. No.: 4,55e-29 Length: 1206
Score: 291.50 Matches: 87
Percent Similarity: 50.41% Conservative: 37
Best Local Similarity: 35.37% Mismatches: 99
Query Match: 21.05% Indels: 23
DB: Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-801 (1-1206)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
DB 514 ATGGAAGGGGCACTGCTGGCCGCGATGCTGGAAGAGACAGCACCGCGGTTCCCGTTCCGTC 573
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39
DB 574 GCCCTGCTGCTTCCGGCGGTGACACCCAGTTGTGCGGGTGAGACGGATATCGGCGCGTAC 633
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
DB 634 CAGTTGCTTGGCGAATCGGTGAGATGCCCGCGGAAGCCTTGACAAAGACCGCCAG 693
QY 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79
DB 694 CTGATCGGCGCTG---GGCTATCCC-----GGTGGTCCGGAATCCGC 732
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAla 99
DB 733 CGCTGCGGAGCGCGGCACTCTGCGCGCTTCCGTTCCCGCGCGCATGACCGCATCGC 792
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
DB 793 CCGGCGCTGAGCTTCAAGCTTCAAGGAGCTTACCTGACAC---ACTGGCAG 849
QY 120 LysGluLysGlnGluGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
DB 850 CGTTGCGTCAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
DB 901 CTGGCGTTCCAGACCGCGGTGTGAGACCTGCTGATCAAGTGCCTGCGCGCTTG--- 957
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGly 179
DB 958 -----AAGAGACCGGCTG-----AAGAAC-----CTGGTATGCGCGCGCT 996
QY 180 ValAlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
DB 997 GTCAAGCCCAACACAGGCGCTGCGAGCGGCTGAAAAGATGCTCGGGAATGAAGGGG 1056
```

[illegible]

Db	552821	GCTCGTTGGATCGGGTGGCCACACTCAATTAGTGGCGTGGATGGTGTAGGAAAAATAT	552762
Qy	40	LeuLeuLeuGlyValSerLeuAsp1Leu1aProGlyAspMetLeuAsp1Val1aAsp	59
	
Db	552761	GAATGATAGGAGAATCTATTGATGATATGCTGTGGCGAAGCCTTGATTAACAACGAAAA	552702
Qy	60	ArgLeuSerLeu1Leu1aHisProGlyCysSerThrMetSerG1Gly1aVal1aLeu1a	79
Db	552701	TTACTTGGACTA---GATTATCCA-----GGTGGCCGGCACACTTTCT	552683
Qy	80	HisLeu1aValysGln1aAsnArgPheHisPheAsp1Leu1aProProLeuHis1aVala	99
Db	552662	CGTTTAGCGGAAAAAGATACCGCAAAATGTTTCACATTTCCACGTCCAAATGACAGATCGT	552603
Qy	100	LyAsnCyAspArgPheSerPheThrGlyLeuGlnHisVal1aThrAsp1aLeu1aMet1a	119
Db	552602	GCAGCGCTGTGATTTTATGTTTCTTGTTTAAAAAATTTGCGCAATACAGTTAATCA	552543
Qy	120	LyAsGlyArgGluGluGly-----1LeuGlySGlyGln1LeuSerSer1a1aAsp	137
Db	552542	GCAATTAAAAACAGGGCGCAACTGATAGACAA-----ACTAAAGCAGAT	552498
Qy	138	1Leu1aVal1aThrValGlnHisPheMet1aCysHis1aLeuVal1aArgThrHisArg1a	157
Db	552497	ATTGTTATGCTTCTTCAAGATCGGTGGATGATCTTGGCC-----	552456
Qy	158	1LeuLeuPheCysLySGlnArgAspLeuLeuProGlnAsnAsn1aVal1aLeuVal1aSer	177
Db	552455	---ATTAAATGTAG--CGTGCAATTGAAGAAAGAACAGCGTATTAACGTTAGATGATGGC	552402
Qy	178	GlyGlyVal1a1aSerAsnPheThr1LeuArgArg1aLeuGlu1LeuThrAsn1a1aThr	197
Db	552401	GGAGGGGAGACGCAATTAATAAACTCCAGAAACGCTGGCACACTTAATGCCAAATTTA	552342
Qy	198	GlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly1LeuMet1a1a	217
Db	552341	GCTGGCGGAAGCTTTTATCTCAACTCACTGATTTGTATGACGATATGCTGATGATGCT	552282
Qy	218	TrpAsnGly1Leu1aArgLeuArg1aGly	227
Db	552281	TACACAGGTTTTTACGTTTAAACAAGGT_552252	
RESULT 11			
US-09-643-990A-1/c			
Sequence 1, Application US/09643990A			
Patent No. 6528289			
GENERAL INFORMATION:			
APPLICANT: Robert D. Fleischmann			
Mark D. Adams			
Owen White			
Hamilton O. Smith			
J. Craig Venter			
TITLE OF INVENTION: The Nucleotide sequence of			
the Haemophilus influenzae Rd Genome, Fragments			
Thereof, and Uses Thereof			
NUMBER OF SEQUENCES: 1			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Human Genome Sciences, Inc.			
STREET: 9410 Key West Avenue			
CITY: Rockville,			
STATE: MD			
COUNTRY: USA			
ZIP: 20850			
COMPUTER READABLE FORM:			
MEDIUM TYPE: 3 1/2 inch diskette			
COMPUTER: Dell Pentium			
OPERATING SYSTEM: MS DOS v6.22			
SOFTWARE: ASCII Text			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/643,990A			
FILING DATE: 23-Aug-2000			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			

APPLICATION NUMBER: 08/487,429
 FILING DATE: 1995-06-07
 APPLICATION NUMBER: 08/426,787
 FILING DATE: 1995-04-21
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB186P1C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-610-5790
 TELEFAX: 310-309-8439
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1830121 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-643-990A-1

Alignment Scores:
 Pred. No.: 1.68e-23 Length: 1830121
 Score: 289.50 Matches: 82
 Percent Similarity: 48.70% Conservative: 30
 Best Local Similarity: 35.65% Mismatches: 95
 Query Match: 20.90% Indels: 23
 Gaps: 7

US-10-649-273-2_COPY_148_414 (1-267) x US-09-643-990A-1 (1-1830121)

1 MetGluAlaHisAlaLeuThrIleArgLeu---ThraNlyValGluPheProPheLeu 19
 552881 ATGGAAGGCGATTACTTGGCCCAATGCTTATGACAACTTCCAGCCACTTCTTTCTT 552822
 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39
 552821 GCCTGTTGGATCGGGTGGCCACACTCAATTAGTGGTGGATGGATGGAAAAATAT 552762
 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
 552761 GAAGTGTAGAGAAATCTATTGAATGATCTCTGCGCAAGCTTGTGAATAAACAAGCAAA 552702
 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlu 79
 552701 TTAATCGACTA---GATTATCCA-----GGTGGCGCGCACTTCT 552663
 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAla 99
 552662 CGTTTAGCGGAAAAAGGTAAGCCCAATCGTTTCACTTCCACGTCGCAATGACAGATCGT 552603
 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
 552602 GCAGGCGCTGATTTAGTTTCTGCTTAAACAATTGCGCGCAAAATACAGTTAATCAA 552543
 120 LysGluLysGlnGluGly-----IleGluLysGlnIleLeuSerSerAlaAlaAsp 137
 552542 GCAATTAAAAACGAGCGCAATCATAGAGCA-----ACTAAAGGCAT 552498
 138 IleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla 157
 552497 ATTGCTTAGCTTTCAGAGATGCGGTGATGACTCTTGCC----- 552456
 158 IleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaAlaIleuValAlaSer 177
 552455 ---ATTAAATGTAG---CGTGCATTGAAAGAAACAGGCTATAAAGCTTTAGTGTTCG 552402
 178 GlyGlyValAlaSerAsnPheTyrIleArgAlaLeuGlnIleLeuThrAsnAlaThr 197
 552401 GAGGGGTGGCGCAATAAATACTCGAGAAACGCTTGGCACTTATGCAAAATTTA 552342
 198 GlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAla 217
 552341 GGTGGCGAAGGTTTATTCCTCAACTCAATTGTGTACAGATAAAGTGCAGATGATGCT 552282

218 TrpAsnGlyIleGluArgLeuArgAlaGly 227
 552281 TACACAGGTTTTCCTTACGTTTAAACAAGCT 552252

RESULT 12
 US-09-902-540-6612
 Sequence 6612, Application US/09902540
 Patent No. 6833447
 GENERAL INFORMATION:
 APPLICANT: Goldman, Barry S.
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Wiegand, Roger C.
 TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
 FILE REFERENCE: 38-10(15849)B
 CURRENT APPLICATION NUMBER: US/09/902,540
 PRIOR FILING DATE: 2001-07-10
 PRIOR APPLICATION NUMBER: 60/217,883
 NUMBER OF SEQ ID NOS: 16825
 SEQ ID NO 6612
 LENGTH: 996
 TYPE: DNA
 ORGANISM: Myxococcus xanthus
 US-09-902-540-6612

Alignment Scores:
 Pred. No.: 6.44e-28 Length: 996
 Score: 282.00 Matches: 86
 Percent Similarity: 48.93% Conservative: 28
 Best Local Similarity: 36.91% Mismatches: 87
 Query Match: 20.36% Indels: 32
 Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-902-540-6612 (1-996)

1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysVal---GluPheProPheLeu 19
 334 CTGAGGGCGCACCTGCTGCGCAATCGCTTGGAGTGGCGCCGAGCGCGCTTCTT 393
 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39
 334 GGGCTGCTGCTTCCGGGGGCGACACAGCTCTTACAGAGTGCAGGCTTACGGGCAATG 453
 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
 454 CGGCTGTTGGGAGACAGCGCGGCGCGCGGCGGAGGAGCATATGACAAAGACCGCTGC 513
 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlu 79
 514 ATCTTCGCGCTG---CCGATATCCG-----GGTGGGCAAGCCCATCGAC 552
 80 HisLeuAlaLysGlnGlyAsn-----ArgPheHisPheAspIleLysProPro 95
 553 CAGTTGGCGGAGGAGGAAACCGGAGGCCATCCGCTT-----CGGCC 597
 96 LeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLys 115
 598 GCGCTCGCGCGGACAACTTGCACCTGCTTCCGCGTTAAG-----ACG 645
 116 IleIleMetLysLysGlnLysGlnGlyIleGluLysGlnIleLeuSerSerAla 135
 646 GCGGCTGCACACAGTGCAGAGGCGCGCGGCGGAGGCGGCGGCGCTG----- 696
 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
 697 GCGATTTGTGCGCTCTTCTTCAAGAGGCGCTGTGCGAGCTGTGCAAGAG----- 750
 156 ArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaAlaIleuVal 175
 751 -----CTGAGGCGCGCGCGCGCGCTTG-----GCCACAAAGCATTTGGTG 792

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 18:03:13 ; Search time 534 Seconds
(without alignments)
2950.230 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALTRLRNKKVEPPFLV.....DISKEVGEASIKVPQKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5378673 seqs, 2950229984 residues
Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1h
-Q=/cg2_1/USPTO.spool/US10649273/runat.14022005.114706.16497/app.query.fasta_1.1429
-DB=Published Applications NA -QFMT=fastap -SUFFIX=mpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS-human40.cdt -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649273 @CGN 1.1.1053 @runat.14022005.114706.16497
-NCPU=6 -ICPU=3 -NO MMAP -LARGECQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=10 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FAPOP=6 -FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA: *
1: /cg2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cg2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cg2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cg2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cg2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cg2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cg2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cg2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cg2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
10: /cg2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
11: /cg2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cg2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq: *
13: /cg2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
14: /cg2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
15: /cg2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
16: /cg2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
17: /cg2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq: *
18: /cg2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq: *
19: /cg2_6/ptodata/2/pubpna/US10F_NEW_PUB.seq: *
20: /cg2_6/ptodata/2/pubpna/US11_NEW_PUB.seq: *
21: /cg2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
22: /cg2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	1416	17	US-10-120-988-177
2	1385	100.0	1526	14	US-10-067-443-23
3	1385	100.0	1526	17	US-10-649-273-23
4	1385	100.0	1526	17	US-10-651-722-23
5	1385	100.0	2197	14	US-10-067-443-1
6	1385	100.0	2197	17	US-10-649-273-1
7	1385	100.0	2197	17	US-10-651-722-1
8	1358.5	98.1	1387	14	US-10-067-443-21
9	1358.5	98.1	1387	17	US-10-649-273-21
10	1358.5	98.1	1387	17	US-10-651-722-21
11	1348	97.3	1245	14	US-10-012-140-6
12	1348	97.3	1820	14	US-10-012-140-4
13	1204	86.9	2208	17	US-10-094-749-400
14	1204	86.9	2890	18	US-10-723-860-7447
15	824.5	59.5	14364	14	US-10-067-443-20
16	824.5	59.5	14364	17	US-10-649-273-20
17	824.5	59.5	14364	17	US-10-651-722-20
18	468	33.8	371	18	US-10-430-201-3118
19	468	33.8	371	18	US-10-430-201-3119
20	422.5	30.5	1917	17	US-10-424-559-66417
21	359	25.9	1628	18	US-10-437-963-11249
22	342	24.7	756	14	US-10-081-051-8
23	342	24.7	4360	14	US-10-081-051-2
24	321.5	23.2	1146	17	US-10-282-122A-14674
25	314.5	22.7	936	17	US-10-282-122A-8315
26	308.5	22.3	1044	17	US-10-282-122A-26972
27	308.5	22.3	94750	17	US-10-672-787-38
28	304.5	22.0	1032	17	US-10-282-122A-31043
29	301.5	21.8	1000	18	US-10-343-561-50
30	291.5	21.0	1026	9	US-09-815-242-7701
31	291.5	21.0	1026	17	US-10-282-122A-30016
32	289.5	20.9	1029	9	US-09-815-242-6546
33	289.5	20.9	1029	17	US-10-282-122A-22020
34	288.5	20.9	1830121	17	US-10-329-670-1
35	288.5	20.9	1830121	18	US-10-158-865-1
36	285	20.6	372	18	US-10-357-930-61088
37	280.5	20.3	1020	9	US-10-282-122A-32254
38	279.5	20.2	1014	17	US-09-815-242-9682
39	279.5	20.2	1014	17	US-10-282-122A-39301
40	279.5	20.2	1014	17	US-10-282-122A-41977
41	277.5	20.0	1007	17	US-10-282-122A-19220
42	277	20.0	927	17	US-10-282-122A-25354
43	275.5	19.9	1014	9	US-10-282-122A-31809
44	275.5	19.9	1014	9	US-09-815-242-6207
45	275.5	19.9	1014	17	US-10-282-122A-20511

ALIGNMENTS

RESULT 1
US-10-120-988-177
; Sequence 177, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyun
; APPLICANT: Wang, Dunhui
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802CON
; CURRENT APPLICATION NUMBER: US/10/120,988
; PRIORITY FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_ft_genes Version 2.0

```
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205) .. (1305)
US-10-120-988-177

Alignment Scores:
Pred. No.: 9,89e-170 Length: 1416
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-120-988-177 (1-1416)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 502 ATGAGAGGCTCATGACCTTACTATTAGGTGACCAATTAAGTAATTTCTTTTATGTT 561
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 562 CTTTGTGATTTCTGAGGTCACCTGCTGTGGCATTTAGTTCAAGAGTTTCAGATTTTCG 621
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 622 CTTCTTGGAAGTCTTTGACATAGCACAGGTGACATGCTTGAACAAGTGGCAAGAAGA 681
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 682 CTTTCTTAAATMAAACATCCAGAGTCTCCACCATGAGTGGGGAAGCCATGAACAAT 741
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 742 TTGGCCAAACAGAGAAATAGATTTCATTTCATGACATCAACCTCCCTTGATCATGCTAAG 801
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 802 AATTGTGATTTTCTTTTACTGGAACCTTCAACCGTTACTGATAAATATATGAAAAAG 861
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 862 GAAAAAGGAGGATGATGAGAAAGGGCAAAATCTGTCTTCAGCAGCAGACATTCCTGCG 921
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 922 ACAGTACGACACACATAGGCATGTCATCTGTGAAAAAGAACACATCGGGCTATATCTGTT 981
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 982 TGTAAAGCAGAGAGCTTGTACTCAAAATATGCAAGTACTGTGATCTGTGTGTGTC 1041
QY 181 AlaSerAsnPheIleIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
DB 1042 GCAGATGAATCTTATATCCGACAGAGCTCTGGAATAATTTTAAACAAACAGACAGTCACT 1101
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleThrAsnGly 220
DB 1102 TTGTTGTGTCTCTCCCTCCAGACTATGACATGATTAATGAGCATTAATATGCAATGAGTGT 1161
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
DB 1162 ATTGAAGACGACAGCTGTGGCTGGCATTTTACATGACATAGAGGACATCGGCTATAGAA 1221
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyGlnAlaSerIleLysVal 260
DB 1222 CCAAAATGTCTCTTGGAGTAGACATATCAAAAGAGTGGAGAGCTTCCATAAAGTA 1281
QY 261 ProGlnLeuLysMetGluIle 267
|||||
```

```
DB 1282 CCACATTAATAAATGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 1.11e-169 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-23 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 1 ATGAGAGGCTCATGACCTTACTATTAGGTGACCAATTAAGTAATTTCTTTTATGTT 60
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 61 CTTTGTGATTTCTGAGGTCACCTGCTGTGGCATTTAGTTCAAGAGTTTCAGATTTTCG 120
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 121 CTTCTTGGAAGTCTTTGACATAGCACAGGTGACATGCTTGAACAAGTGGCAAGAAGA 180
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 181 CTTTCTTAAATMAAACATCCAGAGTCTCCACCATGAGTGGGGAAGCCATGAACAAT 240
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 241 TTGGCCAAACAGAGAAATAGATTTCATTTCATGACATCAACCTCCCTTGATCATGCTAAG 300
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 301 AATTGTGATTTTCTTTTACTGGAACCTTCAACCGTTACTGATAAATATATGAAAAAG 360
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 361 GAAAAAGGAGGATGATGAGAAAGGGCAAAATCTGTCTTCAGCAGCAGACATGCTGCGC 420
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 421 TGTAAAGCAGACACATAGGCATGTCATCTGTGAAAAAGAACACATCGGGCTATATCTGTT 480
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 481 TGTAAAGCAGAGAGACTTGTACTCAAAATATATGCAAGTACTGTGTCATCTGTGTGTGTC 540
QY 181 AlaSerAsnPheIleIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
|||||
```


Db 541 GCAGTAACCTTCTATATCCGAGACTCTGGAAATTTTACAAACGACAGTGCCT 600
Qy LeuLeuCySPProProAryLeuCyThraPaPaNGIYIleMeIlaIaTTPaNGIY 220
Db 601 TTGTTGTGCTCTCTCCGAGACTATGACTGATTAATGACATTATATGACAGAAATGCT 660
Qy 221 ILeGuArGLeUArGAlaGlyLeuGIYIleUuHIsaPILeGIuGIYIleArGYrGIu 240
Db 661 ATTGAAGACACAGCGCTGCTGGGCAATTTTACATGACATGAGAGGATCCGCTATAGA 720
Qy 241 ProLyCySPProLeuGIYAlaSPILeSerLySGluValGIYGIuAlaSerIleLyVal 260
Db 721 CCAAAAGTCCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCCATAAAAGTA 780
Qy 261 ProGInLeuLyMeTGIuIle 267
Db 781 CCACATTTAAATAATGAGATA 801
RESULT 3
US-10-649-273-23
; Sequence 23, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-649-273-23
Alignment Scores:
Pred. No.: 1,11e-169 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 17
US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-23 (1-1526)
Qy 1 MetGIuAlaHIsaIaLeuThrIleArGLeUThraPaPaNGIYIleMeIlaIaTTPaNGIY 20
Db 1 ATGAGGCTATGCACTTACTATTAAGTTGACCAATTAAGTAATTTCTTTTAACTT 60
Qy 21 LeuLeuIleSerGIYGIYHIsCysLeuLeuAlaLeuValGIYGIYAlaSerIleLyVal 40
Db 61 CTTTGAATTTCTGAGGCTCAGTCTGTGTCATTAAGTTCAAGAGCTTTCAGATTTTCTG 120
Qy 41 LeuLeuGIYLySerLeuAlaSPILeAIProGIYAlaSPMeTLeuAlaSPLyValAlaArgArG 60
Db 121 CTTCTTGAAGAGTCTTGGACATGACACAGGTGACATGCTTGAACAGGTGCGAAGAAGA 180
Qy 61 LeuSerLeuIleLySerHIsaPILeGIYCySerThraMetSerGIYGIYAlaIleGIuHIs 80
Db 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACATGAGTGGGAGAAAGCCATAAGACT 240
Qy 81 LeuAlaLySGInGIYAsnArgPheHIsPheAspIleLySPProProLeuHIsHIsaIaLyS 100
Db 241 TTGGCCAAACAAGAGAAATTAATTTTATTTGACATCAAACTCCCTTGCATCATCTAAA 300
Qy 101 AsnCyAspPheSerPheThrGIYLeuGIuHIsValThraSPLyIleIleMeTlyLyS 120

Db 301 AATTGGATTTTCTTTTACTGGACTTCAACGCTTACTGATTAATAATATGAAGAAAAG 360
Qy 121 GIuLySGIuGIYGIYIleGIuLySGIYGIuHIsLeuSerSerIlaIaSPILeAlaIa 140
Db 361 GAAAAAGAGAGATTAATGAAGGGCAATCTGTCTTCAGACAGACATGCTGTCC 420
Qy 141 ThrValGIuHIsEThraMeIlaCySHIsLeuValIleArGYrThraIaArgAlaIleUePhe 160
Db 421 ACAGTACAGACACATATGCAATGCTATCTTGTGAAAAGAACATCGGCTATTTCTGTT 480
Qy 161 CyLySGInArgAspLeuLeuProGInaAsnAlaValIleuValAlaSerGIYGIYAl 180
Db 481 TGTAAAGCAGAGACTGTTTACTCAAAATTAATGACGTAAGTGTGCTCATCTGATGTGTC 540
Qy 181 AlaSerAsnPheThrIleArGArGAlaLeuGIuIleUeThraPaPaNGIYIleArGYrGIu 200
Db 541 GCAGTAACCTTCTATATCCGAGACTCTGGAAATTTTACAAACGACAGTGCCT 600
Qy 201 LeuLeuCySPProProAryLeuCyThraPaPaNGIYIleMeIlaIaTTPaNGIY 220
Db 601 TTGTTGTGCTCTCTCCGAGACTATGACTGATTAATGACATTATTAATGAGAAATGCT 660
Qy 221 ILeGuArGLeUArGAlaGlyLeuGIYIleUuHIsaPILeGIuGIYIleArGYrGIu 240
Db 661 ATTGAAGACACAGCGCTGCTGGGCAATTTTACATGACATGAGAGGATCCGCTATAGA 720
Qy 241 ProLyCySPProLeuGIYAlaSPILeSerLySGluValGIYGIuAlaSerIleLyVal 260
Db 721 CCAAAAGTCCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCCATAAAAGTA 780
Qy 261 ProGInLeuLyMeTGIuIle 267
Db 781 CCACATTTAAATAATGAGATA 801
RESULT 4
US-10-651-722-23
; Sequence 23, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-23
Alignment Scores:
Pred. No.: 1,11e-169 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 17
US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-23 (1-1526)
Qy 1 MetGIuAlaHIsaIaLeuThrIleArGLeUThraPaPaNGIYIleMeIlaIaTTPaNGIY 20
Db 1 ATGAGGCTATGCACTTACTATTAAGTTGACCAATTAAGTAATTTCTTTTAACTT 60

```

QY 21 LeuLeu1eserGlygYhiCysLeuLeu1aleuValGInGlyValSerAspPheLeu 40
| | | | |
DB 61 CTTTGAATTTCTGGAGGTCACGTCTGTGGCATATTGTTCAAGAGCTTCAATTTCTCG 120
| | | | |
QY 41 LeuLeuGlyLysSerLeuAsp11eAlaProGlyAspMetLeuAspLysVal1AaArg 60
| | | | |
DB 121 CTTCTGAAAGTCTTGGACATGACACAGGTGACAGCTTGACAGAGTGGCAAGAGA 180
| | | | |
QY 61 LeuSerLeu1LeuYhiAspProGlyCysSerThrMetSerGlyLysVal1leGluHis 80
| | | | |
DB 181 CTTTCTTAATAAACAATCCAGAGTCTCCACCAATGAGTGGGGAAGGCCATTAACAAT 240
| | | | |
QY 81 LeuAlaLysGInGlyAsnArgPheHisAspAsp11eLysProProLeuHisHisAlaLys 100
| | | | |
DB 241 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 300
| | | | |
QY 101 AsnCysAspPheSerPheThrGlyLeuGInHisVal1ThrAspLys11e1MetLysLys 120
| | | | |
DB 301 AATGTGATTTTCTTTTACTGCACTTCAACACGTTACTGATTAATAATGAAGAAAAG 360
| | | | |
QY 121 GluLysGInGlyLys11eGluLysGlyGIn11eLeuSerSerAla1AaAsp11eAla 140
| | | | |
DB 361 GAAAAAGAGAGAGATTAAGAGAGGAGCAAACTCTGCTTCAGCAGCAGACATTTCTGCC 420
| | | | |
QY 141 ThrValGInHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11eLeuPhe 160
| | | | |
DB 421 ACACTAAGCAGACACATATGCGATGTCATCTTGTAAGAAACAACATCGGCTTATTTCTGTT 480
| | | | |
QY 161 CysLysGInArgAspLeuLeuProGInAsnAsnAlaVal1LeuVal1AaSerGlyVal 180
| | | | |
DB 481 TGTAGAGCAGAGAGCTGTGTACTCAAAATATAGCAGTACTGCTGTGCTGTGCTGTC 540
| | | | |
QY 181 AlaSerAsnPheThr11eArgArgAlaLeuGIn11eLeuThrAsnAla1AaThrGInCysThr 200
| | | | |
DB 541 GCAAGTAACCTTCTATATCCGAGAGCTCTGAAATTTTAAACAAGCACAACAGTGCCT 600
| | | | |
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGly11eMet11eAla1ArgAsnGly 220
| | | | |
DB 601 TTGTTGTCTCTCTCCACAGCTATGCACTGATTAATGACATTATGATTCAGATGGAATGCT 660
| | | | |
QY 221 11eGluArgLeuArgAlaGlyLeuGly11eLeuHisAsp11eGluGly11eArgGlyGIn 240
| | | | |
DB 661 ATTAAGAGACTACAGCTGCTGGCTGGCATTTTACATGACATAGAGGATCCGCTATGAA 720
| | | | |
QY 241 ProLysCysProLeuGlyValAsp11eSerLysGlyVal1GlyGIn1AaSer11eLysVal 260
| | | | |
DB 721 CCAAAATGTCTCTTGGAGTACATATCAAAAGAAAGTTGGAAGCTTCCATTAAGATA 780
| | | | |
QY 261 ProGInLeuLysMetGlu11e 267
| | | | |
DB 781 CCACAATTAAATAATGAGATA 801
| | | | |

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1
Alignment Scores:
Pred. No.: 1,91e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_copy_148_414 (1-267) x US-10-067-443-1 (1-2197)
QY 1 MetGluAlaHisAlaLeuThr11eArgLeuThrAsnLysVal1GluPheProPheLeuVal 20
| | | | |
DB 672 ATGAGAGGCTCATGACATTAATTAATGTTGACCAATTAAGTGAATTTCTTTTATGTT 731
| | | | |
QY 21 LeuLeu1eserGlygYhiCysLeuLeu1aleuValGInGlyValSerAspPheLeu 40
| | | | |
DB 732 CTTTGAATTTCTGGAGGTCACGTCTGTGGCATATTGTTCAAGAGCTTCAATTTCTCG 791
| | | | |
QY 41 LeuLeuGlyLysSerLeuAsp11eAlaProGlyAspMetLeuAspLysVal1AaArg 60
| | | | |
DB 792 CTTCTGAAAGTCTTGGACATGACACAGGTGACATGCTTGACAGAGTGGCAAGAGA 851
| | | | |
QY 61 LeuSerLeu1LeuYhiAspProGlyCysSerThrMetSerGlyLysVal1leGluHis 80
| | | | |
DB 852 CTTTCTTAATAAACAATCCAGAGTCTCCACCAATGAGTGGGGAAGGCCATTAACAAT 911
| | | | |
QY 81 LeuAlaLysGInGlyAsnArgPheHisAspAsp11eLysProProLeuHisHisAlaLys 100
| | | | |
DB 912 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 971
| | | | |
QY 101 AsnCysAspPheSerPheThrGlyLeuGInHisVal1ThrAspLys11e1MetLysLys 120
| | | | |
DB 972 AATGTGATTTTCTTTTACTGCACTTCAACGTTACTGATTAATAATTAAGAAAAG 1031
| | | | |
QY 121 GluLysGInGlyLys11eGluLysGlyGIn11eLeuSerSerAla1AaAsp11eAla 140
| | | | |
DB 1032 GAAAAAGAGAGATTAAGAGAGGAGCAAACTCTGCTTCAGCAGCAGACATTTCTGCC 1091
| | | | |
QY 141 ThrValGInHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11eLeuPhe 160
| | | | |
DB 1092 ACAGTACAGACACATATGCGATGTCATCTTGTAAGAAACAACATCGGCTTATTTCTGTT 1151
| | | | |
QY 161 CysLysGInArgAspLeuLeuProGInAsnAsnAlaVal1LeuVal1AaSerGlyVal 180
| | | | |
DB 1152 TGTAGAGCAGAGAGCTGTGTACTCAAAATTAATGACATGCTGTTGCACTGCTGTC 1211
| | | | |
QY 181 AlaSerAsnPheThr11eArgArgAlaLeuGIn11eLeuThrAsnAla1AaThrGInCysThr 200
| | | | |
DB 1212 GCAAGTAACCTTCTATATCCGAGAGCTCTGAAATTTTAAACAAGCACAACAGTGCCT 1271
| | | | |
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGly11eMet11eAla1ArgAsnGly 220
| | | | |
DB 1272 TTGTTGTCTCTCTCCACAGCTATGCACTGATTAATGAGATTATGATTCAGATGGAATGCT 1331
| | | | |
QY 221 11eGluArgLeuArgAlaGlyLeuGly11eLeuHisAsp11eGluGly11eArgGlyGIn 240
| | | | |
DB 1332 ATTAAGAGACTACAGCTGCTGGCTGGCATTTTACATGACATAGAGGATCCGCTATGAA 1391
| | | | |
QY 241 ProLysCysProLeuGlyValAsp11eSerLysGlyVal1GlyGIn1AaSer11eLysVal 260
| | | | |
DB 1392 CCAAAATGTCTCTTGGAGTACATATCAAAAGAAAGTTGGAAGCTTCCATTAAGATA 1451
| | | | |
QY 261 ProGInLeuLysMetGlu11e 267
| | | | |
DB 1452 CCACAATTAAATAATGAGATA 1472
| | | | |

```

RESULT 6
US-10-649-273-1

RESULT 5
US-10-067-443-1

Sequence 1, Application US/10067443
Publication No. US20030082782A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEIN HIGHLY EXPRESSED IN
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2197
TYPE: DNA
ORGANISM: Homo sapiens

```
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-649-273-1

Alignment Scores:
Pred. No.: 1,91e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-1 (1-2197)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 672 ATGAGGCTCATGCACTACTACTATTAGTTGACCAATAAGTAATTCCTTTTAGTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGATTCTGGAGCTCAGTCTGCTGGTGCATTGATTCAGAGAGTTTCAGATTTCG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 792 CTTCTTGAAAGTCTTTGACATAGCACACAGGTGACATGCTTGACAAAGGTGCAAGAGA 851
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 852 CTTTCTTTAAATAAATCATCCAGAGTCTCCACCATGAGTGGTGGAAAGCCATTAGAACT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 912 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 972 AATGTGATTTTCTTTTAACTGAGCTTCAACACGTTACTGATTAATATATGAAAAAG 1031
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 1032 GAAAAAGAGAAAGTATTGAGAGGGGCAAAATCCGTCTTCAGCGAGACATTCGTGCC 1091
QY 141 ThrValGlnHisIleThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 1092 ACAGTACAGCACACATGCGATGTCATCTTGTAAGAAAGACATCGGGCTATTCTGTTT 1151
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 1152 TCTAAGCAGAGAGCTTGCTGCTCAAAATGAGATGAGTGTGCTGTGCTGTGCTGTC 1211
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
```

```
DB 1212 GCAGTAACTCTATATCCGAGAGCTCTGAAATTTTAACAAACGACACATGCACT 1271
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPrAsnGly 220
DB 1272 TTGTTGTGCTCTCTCCAGACTATGATTAATGGCAATTATGATTGATGGAATGGT 1331
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
DB 1332 ATTTGAAAGACTAGTCTGCTGGCTTGGCATTTTATCAGACATTAAGAGCATTCGCTATGA 1391
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
DB 1392 CCAAAATGCTCTCTTGAGATAGACATATCAAAAGAGTTGAGAGAGCTTCATTAAGA 1451
QY 261 ProGlnLeuLysMetGluIle 267
DB 1452 CCACATTAATAAATGAGATA 1472

RESULT 7
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-651-722-1

Alignment Scores:
Pred. No.: 1,91e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-1 (1-2197)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 672 ATGAGGCTCATGCACTACTACTATTAGTTGACCAATAAGTAATTCCTTTTAGTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGATTCTGGAGCTCAGTCTGCTGGTGCATTGATTCAGAGAGTTTCAGATTTCG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 792 CTTCTTGAAAGTCTTTGACATAGCACACAGGTGACATGCTTGACAAAGGTGCAAGAGA 851
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 852 CTTTCTTTAAATAAATCATCCAGAGTCTCCACCATGAGTGGTGGAAAGCCATTAGAACT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
```

```
Db      912 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTGCATGACTAA 971
Qy      101 AsnCyaspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      972 AATTGTGATTTTCTTTTACTGAGACTTCAACAGTTACTGATTAATATATATGAAAAAG 1031
Qy      121 GltLysGluGluGlyIleGltLysGlyIleLeuSerSerAlaAlaAspIleAlaAla 140
Db      1032 GAAAAAGAGAGAGATTTGAGAGAGGGCAATCTCTCTTCAGCAGCAGACATGCTGCC 1091
Qy      141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db      1092 ACAGTACGACGACACATGAGCATGTCATCTTGAAAAAGAACATCGGGCTATTCTGTTT 1151
Qy      161 CysLysGlnArgAspLeuLeuProGlnIleAsnAlaValLeuValAlaSerGlyGlyVal 180
Db      1152 TGTAGAGCAGAGAGACTTGTTCCTCAAAATATATGAGATCTGTGCACTGTGTGTCTC 1211
Qy      181 AlaSerAsnPhePheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db      1212 GCAAGTAACTTCTATATCCGACAGACTCTGGAATTTTACAAACGCAACAGATGCACT 1271
Qy      201 LeuLeuCysProProPheArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db      1272 TTGTTGTCTCTCTCCAGACTATGCACTGATTAATGCGATTATGATTCATGGAATGCT 1331
Qy      221 IleGluArgLeuAlaGlyIleGluGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
Db      1332 ATTGAAACACATACGCTGCTGCGCATTTTACATGATGAAAGGCATTCGCTATATGA 1391
Qy      241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyAlaSerIleLysVal 260
Db      1392 CCAAAATGCTCTCTTGAGTACATATCAAAAGAGTTGAGAGCTTCAATAAAGTA 1451
Qy      261 ProGlnLeuLysMetGluIle 267
Db      1452 CCACACATTAATAATGAGATA 1472

RESULT 8
US-10-667-443-21
; Sequence 21, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-667-443-21

Alignment Scores:
Pred. No.: 2.74e-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: 14 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-667-443-21 (1-1387)
Qy      1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
|||||
```

```
Db      465 ATGAGAGCTCATGACCTTACTATTAGTTGACCAATAAAGTAAGAAATTCCTTTTATAGTT 524
Qy      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      525 CTTTGAATTTCTGAGAGTACTGTCTGTGGCACTTATGTTCAAGAGATTCAGATTTTCTG 584
Qy      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db      585 CTTCTTGAAATCTCTTTGACATAGCACAGGATGACATCTTGACAAAGGTGGCAAGAA 644
Qy      61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
Db      645 CTTTCTTATATAAATCATCATCAGAGTCTCCACCATGATGTGTGGAAAGCCATAGAGCAT 704
Qy      81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db      705 TTGGCCAAACAGAGAAATGATTTCAATTTTGACATCAACCTCCCTGCACTCAATGCTAA 764
Qy      101 AsnCyaspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      765 AATTGTGATTTTCTTTTACTGACTTCAACAGTTACTGATTAATAATATGAAAAAG 824
Qy      121 GltLysGluGluGlyIle-----GltLysGlyGlnIleLeuSerSerAla 126
Db      825 GAAAAAGAGAGATATATTCTAATTAGTAAAGTTGAACAGATAATATTCCTGATTTG 884
Qy      127 -----GltLysGlyGlnIleLeuSerSerAla 135
Db      885 TGCCTAAAAATAGCTGCTCATTTCTGCAGGTATGAAAGGGGCAAAATCCGTCTTACGA 944
Qy      136 AlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
Db      945 GAGAGATTCCTGCCACAGTACGACACACAAATGCGATTCATTTGTGAAAAAGACACAT 1004
Qy      156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnIleAsnAlaValLeuVal 175
Db      1005 CCGGCTATTCTGTTTGTAAAGAGAGAGACTTGTATCTCCAAATATATGAGTACGTGTT 1064
Qy      176 AlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsn 195
Db      1065 GCATCTGGTGTGTGCGCAAGTAACTTATATCCGAGAGCTCTGGAATTTTAAACAAC 1124
Qy      196 AlaThrGlnCysThrLeuLeuCysProProPheArgLeuCysThrAspAsnGlyIleMet 215
Db      1125 GGAACACAGTGAACCTTGTGTGTCCTCTCCAGACTATGCACTGATTAATGCAATTATG 1184
Qy      216 IleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
Db      1185 ATTGCAATGGAATGATTTGAAAGACTACGTGGCTTGCGCATTTTACATGACATGAA 1244
Qy      236 GltLysArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlyValGlyGlu 255
Db      1245 GGCATCCGCTATGAACCAAAATGTCTCTTGAGTGAATCAATCAAAAGAAATGGAGAA 1304
Qy      256 AlaSerIleLysValProGlnIleLysMetGluIle 267
Db      1305 GCTTCATATAAAGTACCAATTAATAAATGAGATA 1340

RESULT 9
US-10-649-273-21
; Sequence 21, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
```

; PRIOR FILING DATE: 2001-04-10
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 21
 ; LENGTH: 1387
 ; TYPE: DNA
 ; ORGANISM: homo sapiens
 US-10-649-273-21

Alignment Scores:

Pred. No.:	2,74e-166	Length:	1387
Score:	1358.50	Matches:	266
Percent Similarity:	91.10%	Conservative:	0
Best Local Similarity:	91.10%	Mismatches:	1
Query Match:	98.09%	Indels:	25
	17	Gaps:	1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-21 (1-1387)

```

QY 1 MetGluAlaHISAlaLeuThrILeargLeuThraenLysValGluPheProPheLeuVal 20
DB 465 ATGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTAAATTTCTTTTAAATT 524
QY 21 LeuLeuIleSerGlyGlyHISCySLeuLeuAlaLeuValGInGlyValSerAspPheLeu 40
DB 525 CTTTGAATTTCTGAGGTCACGTCTGTTGGCATTAGTTCAAGAGGTTTCAATTTTCTG 584
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 585 CTTCTTGAAAGTCTTTGGACATAGCACAGGTGACATGCTTGACAGGTGGCAAGAGA 644
QY 61 LeuSerLeuIleLysHISProGluCySerThrMetSerGlyLysAlaIleGluHIS 80
DB 645 CTTTCTTTAAATAACATCCAGAGCTCCACATGAGTGGGAAAGCCATTAGAGCAT 704
QY 81 LeuAlaLysGInGlyAsnArgPheHISpHeAspIleLysProProLeuHISHisAlaLys 100
DB 705 TTGGCCAAACAGAGAAATGATTTTCACTTTTGACATCAAACTCCCTTGATCATGCTAA 764
QY 101 AsnCyAspPheSerPheThrGlyLeuGInHISValThrAspLysIleIleMetLysLys 120
DB 765 AATTGTGATTTTCTTTTACTGCACTTCAACAGCTTACTGATTAATATGAAAAAG 824
QY 121 GluLysGInGluGlyIle----- 126
DB 825 GAAAAAGAGAGATATTTCTTAATTAGTAAAGTTGAAACAGATTAATATTCCTGATGG 884
QY 127 -----GluLysGInGInIleLeuSerSerAla 135
DB 885 TGCCATAAAATAGCTGCTCATTTCTGCAAGTATGAAAGGGGCAAAATCCTGTCTTCAACA 944
QY 136 AlaAspIleAlaAlaThrValGInHISThrMetAlaCySHISLeuValLysArgThrHIS 155
DB 945 GCACACATTTGCTGCAACAGTACAGACACAAATGGATGTCATCTGTGAAAAAGAACAT 1004
QY 156 ArgAlaIleLeuPheCySLeuGInArgAspLeuLeuProGInAsnAsnAlaValLeuVal 175
DB 1005 CGGGCTATTCTGTTTGTAGACAGACAGACTGTGTAACCTCAAAATATATGCACTACGTG 1064
QY 176 AlaSerGlyValAlaSerAsnPheThrIleArgAspAlaLeuGInIleLeuThrAsn 195
DB 1065 GCATCTGGTGGTGGCAAGTAACTTCTATATCCGACAGGCTCTGAAATTTTAAACAAAC 1124
QY 196 AlaThrGInCySThrLeuLeuCySProProArgLeuCySThrAspAsnGlyIleMet 215
DB 1125 GCACACACAGTGCATTTCTTGTCTCTCTCCAGACATATGCACTGATATATGCACTTATG 1184
QY 216 IleAlaIlePheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHISAspIleGlu 235
DB 1185 ATTCACAGAAATGGATTAAGAAAGACACUATGTGGCTTGGGCAATTTACATGACATAGAA 1244
QY 236 GluLysLeuArgGlyProLysCySProLeuGlyValAlaAspIleSerLysGluValGlyGlu 255
  
```

```

DB 1245 GCATCCGCTATGAAACCAAAATGCTCTTGGAGTACATATCAAAAGATTGGAGAA 1304
QY 256 AlaSerIleLysValProGInLeuLysMetGluIle 267
DB 1305 GCTTCATTAAGTACCAACATTTAAATATGAGATA 1340
  
```

RESULT 10

US-10-651-722-21

Sequence 21, Application US/10651722

Publication No. US20040048302A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEINASE, MP-1

FILE REFERENCE: D0073 DIV

CURRENT APPLICATION NUMBER: US/10/651,722

CURRENT FILING DATE: 2003-08-29

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 21

LENGTH: 1387

TYPE: DNA

ORGANISM: homo sapiens

US-10-651-722-21

Alignment Scores:

Pred. No.:	2,74e-166	Length:	1387
Score:	1358.50	Matches:	266
Percent Similarity:	91.10%	Conservative:	0
Best Local Similarity:	91.10%	Mismatches:	1
Query Match:	98.09%	Indels:	25
	17	Gaps:	1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-21 (1-1387)

```

QY 1 MetGluAlaHISAlaLeuThrILeargLeuThraenLysValGluPheProPheLeuVal 20
DB 465 ATGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTAAATTTCTTTTAAATT 524
QY 21 LeuLeuIleSerGlyGlyHISCySLeuLeuAlaLeuValGInGlyValSerAspPheLeu 40
DB 525 CTTTGAATTTCTGAGGTCACGTCTGTTGGCATTAGTTCAAGAGGTTTCAATTTTCTG 584
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 585 CTTCTTGAAAGTCTTTGGACATAGCACAGGTGACATGCTTGACAGGTGGCAAGAGA 644
QY 61 LeuSerLeuIleLysHISProGluCySerThrMetSerGlyLysAlaIleGluHIS 80
DB 645 CTTTCTTTAAATAACATCCAGAGCTCCACATGAGTGGGAAAGCCATTAGAGCAT 704
QY 81 LeuAlaLysGInGlyAsnArgPheHISpHeAspIleLysProProLeuHISHisAlaLys 100
DB 705 TTGGCCAAACAGAGAAATGATTTTCACTTTTGACATCAAACTCCCTTGATCATGCTAA 764
QY 101 AsnCyAspPheSerPheThrGlyLeuGInHISValThrAspLysIleIleMetLysLys 120
DB 1065 GCATCTGGTGGTGGCAAGTAACTTCTATATCCGACAGGCTCTGAAATTTTAAACAAAC 1124
QY 121 GluLysGInGluGlyIle----- 126
DB 825 GAAAAAGAGAGATATTTCTTAATTAGTAAAGTTGAAACAGATTAATATTCCTGATGG 884
QY 127 -----GluLysGInGInIleLeuSerSerAla 135
DB 885 TGCCATAAAATAGCTGCTCATTTCTGCAAGTATGAAAGGGGCAAAATCCTGTCTTCAACA 944
QY 136 AlaAspIleAlaAlaThrValGInHISThrMetAlaCySHISLeuValLysArgThrHIS 155
  
```

Db 945 GCAGACATTCGTCACAGTACAGACCAATGCGATGCTTGTGAAAAGACACAT 1004
Qy 156 ArgAlaIleLeuPheCysGlyGlnArgAspLeuProGlnAsnAsnAlaIleVal 175
Db 1005 CGGGCTATTCCTGTTTGTAGAGAGAGACTTGTACCAAAATTAAGCACTACTGCTT 1064
Qy 176 AlAserGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsn 195
Db 1065 GCATTCGTGTGTGTGCGAAGTACTTCTATTCGCCAGAGCTCTGGAATTTTAACAAAC 1124
Qy 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215
Db 1125 GCAACACAGTGCACCTTGTGTGTCTCTCCCAACATATGACATGATATGACATTATG 1184
Qy 216 IleAlaThrPasnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
Db 1185 ATTCGATGAAATGTGATTTGAAGAAGCTACGCTGTGGCTTGGCAATTTTACATGACATAGA 1244
Qy 236 GlyIleArgTyrGlnProGlyCysProLeuGlyValAspIleSerIleGlyValGlyIle 255
Db 1245 GGCATCCGCTTATGACCAAAATGCTCTTGGAGTACATATCAAAAGAGTTGAGAGA 1304
Qy 256 AlAserIleLeuValProGlnLeuLeuMetGluIle 267
Db 1305 GCTTCATTAAGATGACCAATTAATAAGAGATA 1340

RESULT 11

US-10-012-140-6
Sequence 6, Application US/10012140
Publication No. US20030009017A1
GENERAL INFORMATION:
APPLICANT: Leiby, Kevin R.
APPLICANT: Kapeller-Libermann, Rosana
TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
FILE REFERENCE: 381552004900
CURRENT APPLICATION NUMBER: US/10/012,140
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: 60/246,768
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,772
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,185
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 1245
TYPE: DNA
ORGANISM: Homo sapiens
US-10-012-140-6

Alignment Scores:

Pred. No.: 5.45e-165 Length: 1245
Score: 1348.00 Matches: 260
Percent Similarity: 98.50% Conservative: 3
Best Local Similarity: 97.38% Mismatches: 4
Query Match: 97.33% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_COPY 148_414 (1-267) x US-10-012-140-6 (1-1245)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluPheProPheLeuVal 20
Db 442 ATGAGGCTCATGCACTTACTATTTAGTTGACCAATAAGTGAATTTCTTTTATGTT 501
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValAspAspPheLeu 40
Db 502 CTTTGTATTTCTGAGAGTCACTGTCTGTTCGATTAAGTTCAAGAGATTTCATTTTCTG 561

Qy 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValIleArgArg 60
Db 562 CTTCTTGAAAGTCTTTTGACATAGACACAGGTGACATCTTGACAAAGGTGGCAAGAGA 621
Qy 61 LeuSerLeuIleLeuHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
Db 622 CTTTCTTTAATTAACATCCAGAGTCTCCACACAGAGGTGGGAAACCAATAGAAAT 681
Qy 81 LeuAlaIleValGlnGlyAsnAspPheHisPheAspIleLeuProProLeuHisIleAlaIle 100
Db 682 TTGGCCCAACCAAGAAATGATTTCAATTTGACATCAACCTCTTGACATAGCTTAA 741
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetCysIle 120
Db 742 AATTCGATTTTCTTTTACTGACCTTCAACGTTACTAGATTAATAATGAAACAG 801
Qy 121 GlyIleGlyGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 140
Db 802 AAACAAGAGGAAGTATTGAGAAAGGGCAATCTGTCTTCAGCAGACATTCGTGCC 861
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValIleArgThrHisArgAlaIleLeuPhe 160
Db 862 ACAGTACAGACACACATGCGATGTCATCTTGAAAAAGAACATGCGGCTATTCGTT 921
Qy 161 CysIleGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyGlyVal 180
Db 922 TGTAAACAGAGAGACTTGTACTTCAAAATTAATGACATGCTGTTGATCTGTTGTC 981
Qy 181 AlAserAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 982 GCAAGTAACTCTATATATCCGACAGAGCTGGAATTTTAACAAACGACACAGTCACT 1041
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrPasnGly 220
Db 1042 TTGTTGTGTCTCTCCCAAGCTATGCTGATGAATGAGCATTAATGATGATGAGAAATG 1101
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleIleArgTyrGlu 240
Db 1102 ATTGAAAGCTACGTCCTGGCTTGGCAATTTTACATGACATGAAGGCACTCCGCTATGA 1161
Qy 241 ProIleCysProLeuGlyValAspIleSerIleGlyValAlaSerIleLeuVal 260
Db 1162 CCAAAATGCTCTTGAGATGACATATCAAAAGAGTTGAGAAAGCTTCATTAAGAATA 1221
Qy 261 ProGlnLeuLeuMetGluIle 267
Db 1222 CCACATTAATAAGAGATA 1242

RESULT 12

US-10-012-140-4
Sequence 4, Application US/10012140
Publication No. US20030009017A1
GENERAL INFORMATION:
APPLICANT: Leiby, Kevin R.
APPLICANT: Kapeller-Libermann, Rosana
TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
FILE REFERENCE: 381552004900
CURRENT APPLICATION NUMBER: US/10/012,140
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: 60/246,768
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,772
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,185
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1820
TYPE: DNA


```
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (146)...(1390)
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1820)
/ OTHER INFORMATION: n = A,T,C or G
US-10-012-140-4
```

Alignment Scores:

Pred. No.:	9.61e-165	Length:	1820
Score:	1348.00	Matches:	260
Percent Similarity:	98.50%	Conservative:	3
Best Local Similarity:	97.38%	Mismatches:	4
Query Match:	97.33%	Indels:	0
DB:	14	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-012-140-4 (1-1820)

```
QY 1 MetGluAaHsAaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 587 ATGGAGGCTCAGTGAAGTCTTAAAGTTAGGTCACCAATTAAGTAAGATTCTTTTAACTT 646
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuValLeuValGlnGlyValSerAspPheLeu 40
DB 647 CTTTGAATTTCTGAGGCTCAGTCTGTGGCATTAGTTCAAGAGGTTTCAGATTCTCTG 706
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 707 CTTCTTGAAGTCTTGGACATAGCACAGGTGACATGCTTGAACAGGTGGCAAGAGA 766
QY 61 LeuSerLeuIleLysHisAspProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 767 CTTTCTTAATTAATAACATCCAGAGTGTCTCCACATGAGTGTGGGAAAGCCATAGACAT 826
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLys 100
DB 827 TTGGCCAAACAGGAATAGATTATTCATTTTATGACATCAACCTCCCTTCATCATGCTAAA 886
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 887 AATTGATATTTCTTTTACTGAGCTTCAACACGTTACTGATTAATAAATAGAAACACAG 946
QY 121 GluLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 947 AAACAAGAGAGAGGATATTGAGAGAGGCGAAATCTGCTTCAGCAGCAGATTCCTCC 1006
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 1007 ACAGTACAGCACAAATGCGATGTCATCTTGTGAAMAACAACATCGGGCTAATTTCTGTTT 1066
QY 161 CysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAla 180
DB 1067 TGTAAAGAGAGAGCTTGTAACTTCAAAATATATCAGATAGCTGTTGATCTGGTGCTTC 1126
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1127 GCAAGTAACTTCTAATATCCGACAGCTCTGGAATTTTAACAAACGACACAGTGCACT 1186
QY 201 LeuLeuCysProProProAlaGlyLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly 220
DB 1187 TTGTGTGCTCTCTCCACAGCATATGACATGATATATGCAATTAATGATTAATGACATGAT 1246
QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrlle 240
DB 1247 ATTGAAGAAGCTACAGCTGCTGGCTTGGCATTTTATGACATAGAGAGGATCCGCTATATA 1306
QY 241 ProLysCysProLeuGlyValAlaAspIleSerLysGlnValGlyAlaSerIleLysVal 260
DB 1307 CCAAAATGTCTCTCTGAGATAGACATATCAAAAGAGATTGAGAGAGCTTCAATAAAGTA 1366
QY 261 ProGlnLeuLysMetGlnIle 267
```

```
DB 1367 CCACATTAATAATGAGATA 1387
```

RESULT 13

US-10-094-749-400

Sequence 400, Application US/10094749

Publication No. US20030219741A1

GENERAL INFORMATION:

```
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHITO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARA, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
FILE REFERENCE: 08435/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 400
LENGTH: 2208
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-400
```

Alignment Scores:

Pred. No.:	7.67e-146	Length:	2208
Score:	1204.00 <td>Matches:</td> <td>239</td>	Matches:	239
Percent Similarity:	89.51%	Conservative:	0
Best Local Similarity:	86.93%	Mismatches:	4
Query Match:	86.93%	Indels:	24
DB:	17	Gaps:	1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-094-749-400 (1-2208)

```
QY 1 MetGluAaHsAaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 765 ATGGAGGCTCAGTGAAGTCTTAAAGTTAGGTCACCAATTAAGTAAGATTCTTTTAACTT 844
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuValLeuValGlnGlyValSerAspPheLeu 40
DB 845 CTTTGAATTTCTGAGGCTCAGTCTGTGGCATTAGTTCAAGAGGTTTCAGATTCTCTG 904
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 905 CTTCTTGAAGTCTTGGACATAGCACAGGTGACATGCTTGAACAGGTGGCAAGAGA 964
QY 61 LeuSerLeuIleLysHisAspProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 965 CTTCTTGAATTAATAACATCCAGAGTCTTCACATAGAGTGGGAAAGCCATAGACAT 1024
QY 81 LeuAlaLysGlnGlyValAlaAspIleSerLysGlnValGlyAlaSerIleLysVal 100
DB 1025 TTGGCCAAACAGGAATAGATTATTCATTTTATGACATCAACCTCCCTTCATCATGCTAAA 1084
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
```


|||||
Db 1085 AATTGATGATTTCTCTTACTGAGACTTCACACGTTATGATTAATAATATGAAAAAG 1144
Qy 121 G|U|U|S|G|U|G|U|G|Y|I|E|G|U|U|S|G|Y|G|I|N|I|E|U|S|E|S|E|S|E|A|A|A|S|P|I|E|A|A|A| 140
Db 1145 GAAAAAGAGGAAGGATATYGAGAAAGGGGCAAAATCTGTCTTCAGCAGCAGACATTCCTGCC 1204
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleuPhe 160
Db 1205 ACAATTACGACACACAAATGGCAGATGTCATCTTGTAAGAAAGACATCGGGCTATTCGTTT 1264
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
Db 1265 TGTAAAGCAGAGAGACTTGTATCTCAAAATATGACAGTACTGTTGCATCTGTGTGTC 1324
Qy 181 AlaserAsnPhetYrIleArgArgAlaLeuGluIleuThrAsnAlaThrGlnCysThr 200
Db 1325 GCAGTAACTTCTGTATCCGACGAGCTCTGGAATTTTAAACAAACGCAACAGTGCAC 1384
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
Db 1385 TTGTTGTGTCCTCTCCACGACTATGACATGATTAATGGCATTATGATTGCA----- 1435
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleuHisAspIleGluGlyIleArgTyrglu 240
Db 1435 ----- 1435
Qy 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
Db 1436 ---TGATGTCTCTTGGAGTACATATCAAAAGAGTTGGAGAGCTTCCTCAATAAGTA 1492
Qy 261 ProGlnLeuLysMetGluIle 267
Db 1493 CCACAATTAAAAATGAGAGATA 1513

RESULT 14

US-10-723-860-7447
; Sequence 7447, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Azile, Natacha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: US/10/723, 860
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7447
; LENGTH: 2890
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (646)..(657)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-7447

Alignment Scores:

Pred. No.: 1,15e-145 Length: 2890
Score: 1204.00 Matches: 239
Percent Similarity: 89.51% Conservative: 0
Best Local Similarity: 89.51% Mismatches: 4
Query Match: 86.93% Indels: 24
DB: 18 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-723-860-7447 (1-2890)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20

|||||
Db 1442 ATGAGGCGTACGCACTTACTATGTTGACCAATAAGTGAATTCCTTTTAGTT 1501
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspLeu 40
Db 1502 CTTTGAATTTCTGGAGGTCACTGTCTGTGGCATTAAGTTCAAGAGATTTTCAGATTTTCG 1561
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 1562 CTTCTTGAAAGTCTTTGACATAGCACCGGATGATCTTGACAGAGGTGGCAAGTA 1621
Qy 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
Db 1622 CTTCTTTATTAATAACATCCAGAGTCTTCCACATGAGTGGTGGAAACCATAGAACAT 1681
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 1682 TTGGCCAAACAAAGAAATGATTTGATTTGACATCAACCTCCCTTGATCATGCTATA 1741
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 1742 AATTGATTTTCTTTTACTGACTTCAACAGTTACTGATTAATAATGAATAAG 1801
Qy 121 G|U|U|S|G|U|G|U|G|Y|I|E|G|U|U|S|G|Y|G|I|N|I|E|U|S|E|S|E|S|E|A|A|A|S|P|I|E|A|A|A| 140
Db 1802 GAAAAAGAGGAAGGATTTGAGAAAGGGGCAAAATCTGTCTTCAGCAGCAGACATTCGTGCC 1861
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleuPhe 160
Db 1862 ACAGTACAGCACACATAGTCATCTTGAAAGAAAGAACATCGGGCTATTCGTTT 1921
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
Db 1922 TGTAAAGCAGAGACTTGTATCTCAAAATATGACATGCTGTGCATCTGTGTGTC 1981
Qy 181 AlaserAsnPhetYrIleArgArgAlaLeuGluIleuThrAsnAlaThrGlnCysThr 200
Db 1982 GCAGTAACTTCTGTATCCGACGAGCTCTGGAATTTTAAACAAACGCAACAGTGCAC 2041
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
Db 2042 TTGTTGTGTCCTCTCCACGACTATGACATGATTAATGCAATTATGATTGCA----- 2092
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleuHisAspIleGluGlyIleArgTyrglu 240
Db 2092 ----- 2092
Qy 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
Db 2093 ---TGATGTCTCTTGGAGTACATATCAAAAGAGTTGGAGAGCTTCCTCAATAAGTA 2149
Qy 261 ProGlnLeuLysMetGluIle 267
Db 2150 CCACAATTAAAAATGAGAGATA 2170

RESULT 15

US-10-067-443-20
; Sequence 20, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D00073 NP
; CURRENT FILING DATE: US/10/067,443
; PRIOR APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20

LENGTH: 14364
TYPE: DNA
ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:
Pred. No.: 3,84e-95
Score: 824.50
Percent Similarity: 26.73%
Best Local Similarity: 26.73%
Query Match: 59.53%
Length: 14364
Matches: 240
Conservative: 0
Mismatches: 2
Indels: 657
Gaps: 4

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-20 (1-14364)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 10845 ATGAGAGCTCAATGACCTTATAGGTGACCAATAAAGTAAGATTCTTTTATAGTT 10904
QY 21 LeuLeuLysSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 10905 CTTTGTATTTCTGGAGGTCACGTGCTGTGGCATTTAGTTCAAGAGGTTTCAGATTTCCTG 10964
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysVal----- 57
DB 10965 CTTTCATGAAAAGCTTTTGGACATAGCAACAGGTGACATGCTTGACAAAGT-AAATTAAGAA 11023
QY 57 ----- 57
DB 11024 TTAATTTCTCCATTCTTTTGTATGTGTGCATTTCACTAAAGTAGCAATAGATGTGC 11083
QY 57 ----- 57
DB 11084 TACCACCATTCACCTAAATATTCTGAAATTTTACTAGTAACTGAAAAAATTACAT 11143
QY 57 ----- 57
DB 11144 ATGTGAGAAAAAATAGAAAAGTAGTACACAAATTTTATATTCTTACCTTTCTTAATA 11203
QY 57 ----- 57
DB 11204 AAATGTAGAGGTTTCATATCTGTATACATAAAGCTGAAATAGTTTGACATACAGTTATG 11263
QY 57 ----- 57
DB 11264 TATTTTGCAAATATATGTATGTGAAGAAAGTGTCTGTAACTAATACACTGCAAAAAA 11323
QY 57 ----- 57
DB 11324 GGTAAATAAGAGATATATATAGATTTAACATAAGACATTAAAGATGCAATGACAGAA 11383
QY 57 ----- 57
DB 11384 TTAATACACAAATTACTTACACACAGACAGAGGTCCCGCCCAACCCCGTTTGTTTAG 11443
QY 57 ----- 57
DB 11444 AATATACAGAGGCTACTGCTCATATATAGAAAACAAACAAACAAACAAACACTG 11503
QY 57 ----- 57
DB 11504 CTTCCACAGCTGAATATATAGAAATATAGCAAGATTCTTATATATGACCTTCATCAT 11563
QY 57 ----- 57
DB 11564 AAGCAGTATTGTCAACTTCAAGCCCAATTTTCAACCAATAGAAAGAGAAACATAGACAG 11623
QY 57 ----- 57
DB 11624 GGGCAGATTTGGCCCTTATTTGTTGGGTTCATATAGAAACAGGGTGTCTGCTTACC 11683
QY 57 ----- 57

DB 11684 TGAATATCAGCTATAGTCTATATTTGCCAAAGTATAGCATGTTTATTCATTCAGGGGTT 11743
QY 57 ----- 57
DB 11744 TTTTGTTTGTAGTATTTCAATTATTTCTTTGACCTTTTGTGTTTACAGATATTT 11803
QY 58 ----- 11803
DB 11804 AATTTTATGACTTAAAAATATATGTTTCTTGTATAGTGGCAAGAGACTTTCTTTAATA 11863
QY 65 LysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGlnHisLeuValGln 84
DB 11864 AAACATCCAGAGTGTCCACCATGAGTGGGAAAGCCATAGAACATTTGGCCAAACAA 11923
QY 85 GlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPhe 104
DB 11924 GGAATATAGTTTCATTTTGAACATCAAACTCCCTTCATCATGCTTAAAAATGTGATTTT 11983
QY 105 SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlyLys 124
DB 11984 TCTTTTACTGACCTTCAACACCTTACTGATTAATATATGAAAAAGAAAAAGAGAA 12043
QY 124 ----- 124
DB 12044 GGTATATTTCTAATTAGTAAAGTTGAACAGATTAATATTTCTGATGTGCTTAAATA 12103
QY 125 ----- 12103
DB 12104 GCTGCTCATTTCTGACAGGTATATGAGAAAGGGCAAACTCTGCTTCGACAGCAACATTC 12163
QY 139 AlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLe 159
DB 12164 TGCCACAGTACAGCACACATGCAATGCAATGCTTGTGAAAAAGAACACATGGGCTATTTCT 12223
QY 159 uPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeu----- 174
DB 12224 GTTTGTAGAGAGAGAGACTTGTACTCAAAATATAGCAGTACTGTAAAGTTTATTC 12283
QY 174 ----- 174
DB 12284 ATTTATAGTATATAGTTACACTTTCGAAATATGTACTTTTCCCAAGACTTGACCTTG 12343
QY 174 ----- 174
DB 12344 TGTTAGATGAACAGACTTTATGCTTATGCTTATGCTGACCTGACAGTAAATATAGAG 12403
QY 174 ----- 174
DB 12404 GATAGAAAGACTAACAGCATTTCTTGTACTAGTTGTGTAGCTTATAGGACAGCTGTA 12463
QY 174 ----- 174
DB 12464 TAGCTTATAGGACATATAGTAAATTTGATCTTCTGTGATTTAAAGAGGCTT 12523
QY 174 ----- 174
DB 12524 ACAATTAAGAAAGTAATGACAGTAACTGCTATCACTATTTTAAAGAAATAGGTGATTT 12583
QY 174 ----- 174
DB 12584 CCTTATCTTTGATGAATATCCCTTGTGTTTGTGTTTAAATATAGCAGTCAATTT 12643
QY 174 ----- 174
DB 12644 AGCAGTGGAGGTGATTTCAACTTGTGACACTAATGTTGATTAAGTTGATATATC 12703
QY 174 ----- 174
DB 12704 CACTATATTTGACACCAAAATCCCTTAAATGTGCTTAAAGCTTGAACAACATCT 12763
QY 174 ----- 174
DB 12764 GTTAACTGTATCTTAACTTATTTCAATTTAAATATATTAACATAAGTGGGAAATGTT 12823

```
QY      174 ----- 174
DB      12824 TAAATGTAAGTAATTCATAGATGGAATTTTACATGATATCAAGAAATATATTTTTCAGA 12883
QY      174 ----- 174
DB      12884 GTTATGTAGTAAATGCAAAATATATAAATTTCAAGGTCTAAATAGTACTATAGA 12943
QY      174 ----- 174
DB      12944 TTGAAATTAATATTAATAATTAATTTAGATGAAGGTGGAAGAAATATACAAAATGCT 13003
QY      174 ----- 174
DB      13004 AGTAATGTTGTATGCTATTAAGAAATTAATTAATTTTCTTCCAAATTTTATTATAC 13063
QY      174 ----- 174
DB      13064 ATGATATATGTCATCTGCCATTAACCATCTCAAAATGGATAGTTATATTGTTAANG 13123
QY      174 ----- 174
DB      13124 CTGATATTTTCTCCAGGTTAAATTAAGCAGCTGGTTCATATCCATATATGATAGTTATT 13183
QY      175 -----ValAlaSerGlyGlyValAlaSerAspPheTyrIleAr 187
DB      13184 TTGGTTTTCATCAATTCCTTCAGGTTGCACTGGTGGTGTGCGAGTAACCTTCTATATATCCG 13243
QY      187 GATGAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAr 207
DB      13244 CAGAGCTCTGGAATTTTAACAAGCAACAGTGCACTTGTGTGTCTCTCCCG 13303
QY      207 gLeuCysThrAspAsnGlyIleMetIleIleArP----- 218
DB      13304 ACTATGCACTGATATGCAATTATGATTCATGTAGCAACAGGATATACGTCTCAC 13363
QY      218 ----- 218
DB      13364 TCATTACTATGTAATAATTAATTCGCATTTTATCATACTAAAGCTTCTCTTCAGATCT 13423
QY      219 -----AsnGlyIleGluArgLeuArgA 226
DB      13424 TGGAGCTATGATTTTATTTTAATGCTTCTTATTTAGGAATGATTTGAAAGACTACGTG 13483
QY      226 laGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLys 242
DB      13484 CTGGCTTGGGCAATTTTACATGACATAGAAAGCATCCGCTATGAACCAAG 13533
```

Search completed: February 17, 2005, 01:27:09
Job time : 559 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:38:06 ; Search time 12.1891 Seconds
(without alignments)
2107.605 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALTIRLTNKVEPFLV.....DISKEVGASIKVPQKMEI 267

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	409.5	29.6	463	E84888	probable O-sialogl
2	338	24.4	365	D97677	O-sialoglycoprotei
3	338	24.4	366	D97677	probable O-sialogl
4	338	24.4	387	E71711	probable O-sialogl
5	334.5	24.2	359	AB3274	O-sialoglycoprotei
6	329.5	23.8	344	E97707	O-sialoglycoprotei
7	326.5	23.6	367	F87257	peptidase M22 fami
8	291.5	21.0	341	H83572	O-sialoglycoprotei
9	290.5	21.0	335	G70369	O-sialoglycoprotei
10	289.5	20.9	342	H64074	O-sialoglycoprotei
11	279.5	20.2	337	AI0079	probable glycoprot
12	279.5	20.2	337	AG0892	probable glycoprot
13	278.5	20.1	337	C91122	probable O-sialogl
14	278.5	20.1	337	H85967	probable O-sialogl
15	276	19.9	421	T18825	hypothetical prote
16	275.5	19.9	337	Q0E8R6	O-sialoglycoprotei
17	275	19.9	412	T40889	probable proteinas
18	270.5	19.5	325	A38108	O-sialoglycoprotei
19	265.5	19.2	354	C81040	O-sialoglycoprotei
20	263.5	19.0	354	C81986	probable O-sialogl
21	263	19.0	348	D82807	O-sialoglycoprotei
22	256	18.5	346	H70195	O-sialoglycoprotei
23	253	18.3	336	C97888	O-sialoglycoprotei
24	250.5	18.1	346	F69786	glycoprotein endop
25	250	18.1	336	B95015	glycoprotein endop
26	244.5	17.7	343	D63718	hypothetical prote
27	243	17.5	337	G72411	hypothetical prote
28	242.5	17.5	348	S75548	O-sialoglycoprotei
29	241.5	17.4	346	AF1820	O-sialoglycoprotei

30	239.5	17.3	344	2	AC1334	glycoprotein endop
31	237.5	17.1	340	2	B97011	probable O-sialogl
32	236.5	17.1	336	2	E84936	O-sialoglycoprotei
33	233	16.8	323	2	G69388	O-sialoglycoprotei
34	231	16.7	344	2	AB1705	glycoprotein endop
35	229.5	16.6	346	2	G86661	O-sialoglycoprotei
36	222.5	16.1	324	2	F75029	O-sialoglycoprotei
37	220	15.9	338	2	A71545	probable O-sialogl
38	219.5	15.8	534	2	H69056	O-sialoglycoprotei
39	219	15.8	335	2	E81278	probable glycoprot
40	218	15.7	344	2	H72106	O-sialoglycoprotei
41	218	15.7	344	2	B86515	O-sialoglycoprotei
42	217	15.7	341	2	G89996	hypothetical prote
43	216	15.6	407	2	S50740	OR17 protein - yea
44	215.5	15.6	324	2	C71215	O-sialoglycoprotei
45	215.5	15.6	344	2	H70737	probable O-sialogl

ALIGNMENTS

RESULT 1
E84888
probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #ext_change 02-Jun-2003
C.Accession: E84888
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A.Reference number: A84420; MUID:20083487; PMID:10617197
A.Accession: E84888
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-463 <STO>
A.Cross-references: GB:AE002093; NID:g2583127; PIDN:AA82636.1; GSPDB:GN00139
C.Genetics:
A.Gene: At2g45270
A.Map position: 2
C.Superfamily: O-sialoglycoprotein endopeptidase

Query Match	Score	29.6%	Score	409.5	ID	Length	463
Best Local Similarity	36.6%	Pred. No.	5.1e-29				
Matches	100	Conservative	37	Mismatches	91	Indels	45
							Gaps 6
QY	1	MEAHALTIRLT-NKVEPFLVLLISGHCIALAVGVSDPFLGSKSDIAPGMDLPKVAR	59				
DB	192	MEAHALTIRLT-NKVEPFLVLLISGHCIALAVGVSDPFLGSKSDIAPGMDLPKVAR	251				
QY	60	RLSLIKHEPCTSMGSKAIEHLAKQGNRHPDIKPLRHAKUCDFSPGICQHTYKTIK	119				
DB	252	WLGDMH-----RSGGPAVEBELALSGDAKSVENPMKXKDCNFSYAGIKQVRLATA	306				
QY	120	KEKEGIEKQGLSSADIAATVQNTMACHLYKTRHAILFCQKQDLQONNAVVASGG	179				
DB	307	KE-----IRNRADIASFORVAVALHEKCEKRAIDWLE--LSPSIKHVWISG	353				
QY	180	VASNFYIRALEILTNNAQCTLLCPPRRLCTNGMIANGIERLAPAGGLHIDEGIRY	239				
DB	354	VASNFYIRALEILTNNAQCTLLCPPRRLCTNGMIANGIERLAPAGGLHIDEGIRY	403				
QY	240	B-----PKCPGLVDISKVGEA	256				
DB	404	DEPPRATPEBDVVDLRPRMPLGSEYAKGRSA	436				
RESULT 2							
AB2902		O-sialoglycoprotein endopeptidase GCP (imported) - Agrobacterium tumefaciens (strain C58,					
C.Species:		Agrobacterium tumefaciens					
C.Date:		11-Jan-2002 #sequence_revision 11-Jan-2002 #ext_change 18-Nov-2002					

C:Accession: AB2902
R:Wood, D.W.; Senubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, M.; Grant, C.; Guentherer, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
Karp, P.; Romero, F.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2902
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <KIR>
A:Cross-references: GB:AE008688; PID:AL43632.1; PID:g17741154; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: gcp
A:Map position: circular chromosome
C:Superfamily: O-staloglycoprotein endopeptidase

Query Match 24.4%; Score 338; DB 2; Length 365;
Best Local Similarity 34.9%; Pred. No. 1.1e-22;
Matches 90; Conservative 40; Mismatches 98; Indels 30; Gaps 7;

OY 1 MEAAATLRLTNKVEPFLVLLISGGCHLTALVQVSDFLLGSKLDIAPDMDKVARR 60
| | | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 120 LEHALTARLTDIGSFPYLMLVSGGHQTQLVAVGVGEYERWGTITDDALGEARDKTAKL 179
61 LSLIKHEPCSTMSGGKAIEHLAKOGNRPHPDIKEPLHHANXCDISFTGL----QHVTDKI 116
Oy 180 LGT-PYP-----GGPAVENNAAGDPDRFPLPRPMVGEARLDPSFGSLKTAVAQAATAI 232
Db 117 IMKKEKEEGLEKGQILSSADIATTVQHTMACHVKETHRALIFCKQRDLIPONNA--VL 174
OY 233 APLESD-----IADICASFOKVASRTLKDIRIGRLAFKYE--FPHINGEPAL 279
Db 175 VASGSVANSPFIIRBALRILMTAQCTLCACPPRICPTNGIMIANGIERLAGIGLIHDI 234
| : | | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 280 VWAGSVANAODEIRDTLALCPTHGFVFVAPHRLCTNNAMAIAWAGERNAEG---ROA 335
OY 235 EGRIYPEKCPDGVDYSKE 252
| : | : | : | | : | : | : | : | : | : | : | : | : | : | : | :
Db 336 DALEVAPERSRWPIDGSAB 353

RESULT 3
D97677
probable o-sialoglycoprotein endopeptidase (glycoproteinas) [imported] - Agrobacterium C/Spectes: Agrobacterium tunefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C/Accession: D97677
R/Goodner, B.; Hinkley, G.; Gatungu, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, W.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B., Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens number: A97359; PMID:117433194
A/Accession: D97677
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-366 <KUR>
A/Cross-references: GB:AEO07869; PID:NARK8373.1; PID:g15157858; GSPDB:GN00169
C/Genetics:
A:Gene: AGR_C_4806
A:Map position: circular chromosome
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.4%; Score 338; DB 2; Length 366;
Best Local Similarity 34.9%; Pred. No. 1.1e-22;
Matches 90; Conservative 40; Mismatches 98; Indels 30; Gaps 7;

OY 1 MEAAATLRLTNKVEPFLVLLISGGCHLTALVQVSDFLLGSKLDIAPDMDKVARR 60
| | | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 121 LEHALTARLTDIGSFPYLMLVSGGHQTQLVAVGVGEYERWGTITDDALGEARDKTAKL 180

Oy	61	ISLIIHPCSTMSGKAIETHLAKOGRHFHDJLPLPHANOCPSFTGL----	QHATDKI	116
Oy	181	LGL-PYP-----GGPAVENAAAKGDPDRFLPRLPVPVGEARLDFSGLKTARQAATAI		233
Oy	117	IMCKEKEGIEKGQILSSAADI AATVQHTMACHLVERTHRAILFCCKORDLLPONNA--	VL	174
Db	234	APLSBQD-----IADI CASFOKAVSRTLKDRIGRLGRLPKVE--	FPHINGEPAL	280
Oy	175	VASGGVANSFYTRALAEITNATQCTLCPPRLCTDNGIMTAWNGIERLRAGIGILHDI		234
Db	281	VVAGGVANAOEIRQTLQALCDTHGRFVAPPHRLCTDNAMIAMAGLERMAEG----	ROA	336
Oy	235	EGIRYEPKCPGLGVDISKB		252
Db	337	DALEVAAPRSRWPLDGSAB		354

```

RESULT 4
E71711
probable O-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
C:Accession: E71711
R:Andersson, S.G.R.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9623693
A:Accession: E71711
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-387 <AND>
A:Cross-references: UNIPROT:Q9ZE8; GB:AJ235270; GB:AJ235269; NID:33860572; PIDN:CAAI4506
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: gcp;
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match      24.4%; Score 338; DB 2; Length 387;
Best Local Similarity 32.5%; Pred. No. 1,1e-22;
Matches 87; Conservative 38; Mismatches 83; Indels 60; Gaps 6;

OY 1 MEAHALTRLTNKEVPEPLVLTISGHCILALVOGVSDTLLEKSLDIAPGMLDKVARR 60
    ||||| ||| : : : ||||| ||||| : : : ||||| : : : ||||| :
Db 113 LEQHATLRLTNDNISGYPIYLLLASGHCQFVAVLGKTKKIIGTTIDDAVGEFTDKAKM 172
    ||||| ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
OY 61 LSLIHPROCSYMSGKALIEHLAKQGNRFHFDIKPLIHLAKNCDPSTGLCHVTDKITMK- 119
    : : : : : ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
Db 173 LNL-----SPGGSEIEIKRAKLNPHKYPKPKPKPIINSGNCMSFSLKTAVRTLLMNL 225
    ||||| ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
OY 120 KEKEGIEKQGLSSAADIAATVQHTMACHLVYKTRATILFCQK----- 163
    ||||| ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
Db 226 KEVNDV-----INDIASFOFTIGAILSSKMDAIRLYQIINDYEDINHPKLN 277
    ||||| ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
OY 164 -----RD-----LLPON-----NAVIVASGVASNFYIRRALIEILTN 195
    : : : : : ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
Db 278 LKSPKDFNMKPELCITRPKRYRIHQNSYRSNLNDTIVIGVAVANKYLQETILSDCTR 337
    ||||| ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
OY 196 ATQCTILCPPEPLCTDNGIMIMANGIER 223
    ||||| ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :
Db 338 PYGYRLIAPPMHLCTDNAMIAIYAGLER 365
    ||||| ||||| : : : ||||| ||||| : : : ||||| : : : ||||| :

RESULT 5
AB3274
O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Brucella melitensis (strain
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AB3274
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.;
Mazur, M.; Goltzman, B.; Selciov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesac
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

```

A:Reference number: AD3252; PMID:11756688
A:Accession: AB3274
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-359 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51357.1; PID:g17982056; GSPDB:GN00150
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME10175
A:Map position: 1
C:Superfamily: O-sialoglycoprotein endopeptidase
C:Keywords: hydrolyase; metalloproteinase

Query Match 24.2%; Score 334.5; DB 2; Length 359;
Best Local Similarity 37.6%; Pred. No. 2.2e-22;
Matches 86; Conservative 33; Mismatches 83; Indels 27; Gaps 6;

OY 1 MEAHATITRLTNKVEPFLVLLISGGHCLALVGVSPDLGLGKSLDIAPGMLDKVAR 60
Db LEGHATITRLTDLGPFYLLLVSGHTQWLVRGIGDERLGTITIDDLGSAFDTKAK 175
OY 61 LSLIKAPECSWGSCKAIEHLAKQGNRFHPDIKPLLHAKNCDFSTGLQHTVDKILMK 115
Db LGL-FVP-----GGPAVERNALQDQKRPALPRLLKGEARLDFSPSGLKTAVRGATTEL 228
OY 116 IIMKEKEGIEKGQILLSSAADIAATVQHTMACHVKTTHRAILFCQKDLFQ-NNAVL 174
Db VRLTLD-----DVTDICASFQAAVADTISDRGRSLERKTE--PDCATPSSL 274
OY 175 VASGVAANFYIRALEILTNATQCTLLCPPRCLTDNGIMAMNGIER 223
Db VVAGVAANNTLRALLENLCTRHGPAFIAPPLNLTCTDNAMAMTAMAGAER 323

RESULT 6
E97707
O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Rickettsia conorii (strain C/Species: Rickettsia conorii
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C/Accession: E97707
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.B.; Barbe, V.; Samson, D.; Rickettsia conorii
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: E97707
A:Status: preliminary
A:Accession: E97707
A:Molecule type: DNA
A:Residues: 1-344 <KUR>
A:Cross-references: GB:AE006914; PIDN:RAL02599.1; PID:g15619097; GSPDB:GN00173
C:Genetics:
A:Gene: gcp
C:Superfamily: O-sialoglycoprotein endopeptidase
C:Keywords: hydrolyase; metalloproteinase

Query Match 23.8%; Score 329.5; DB 2; Length 344;
Best Local Similarity 36.0%; Pred. No. 5.8e-22;
Matches 81; Conservative 40; Mismatches 85; Indels 19; Gaps 5;

OY 1 MEAHATITRLTNKVEPFLVLLISGGHCLALVGVSPDLGLGKSLDIAPGMLDKVAR 60
Db LEGHATITRLTDLGPFYLLLVSGHCLQVAVAGLIGKTKLSTIDDAVGEAFDKVAKM 172
OY 61 LSLIKAPECSWGSCKAIEHLAKQGNRFHPDIKPLLHAKNCDFSTGLQHTVDKILMK 119
Db LNL-----AFQCGPEIRERAKGDPHKYKPRKIIINSGNCMSFSGLKTAVRITLML 225
OY 120 KEKEGIEKGQILLSSAADIAATVQHTMACHVKTTHRAILFCQKDLFQ-NNAVL 178
Db KEINTV-----INDIAASFQFTIGELISKQDAIRAVEQJTNPDFKKN--IVIAG 275
OY 179 GVAANFYIRALEILTNATQCTLLCPPRCLTDNGIMAMNGIER 223
Db GVAANNTLRALLENLCTRHGPAFIAPPLNLTCTDNAMAMTAMAGAER 320

RESULT 7

P87257
peptidase M22 family protein [imported] - Caulobacter crescentus
C|Species: Caulobacter crescentus
C|Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C|Accession: P87257
R.Nieman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.K.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonel
n, U.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A|Title: Complete Genome Sequence of Caulobacter crescentus.
A|Reference number: AB7249; MUID:21173698; PMID:11259647
A|Accession: P87257
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-367 <STO>
A|Cross-references: GB:AE005673; NID:g13421168; PIDN:AAK22058.1; GSPDB:GN00148
C|Genetics:
A|Gene: CC0071
C|Superfamily: O-sialoglycoprotein endopeptidase

Query Match 23.6%; Score 326.5; DB 2; Length 367;
Best Local Similarity 36.3%; Pred. No. 1, 2e-21;
Matches 89; Conservative 36; Mismatches 99; Indels 21; Gaps 6;

OY 1 MEAHALTRLNKVERPPLVLLISGGHCLLAVQGVSDFLILGKSLDIAPGMIDRYARR 60
 ::: :: : ::::: ||| |::| : ::| : ::|
Db 124 LEGHAVSARLGADIVAPPFLILLVSGHCOLLEVSQVACRKRLTITDDAAGEAFDKIAXS 183
 ::: ::| : ::||| |::| |::| : ::| : ::|
OY 61 LSIIHGPECSMTSGSKATIEHLAKCGNRHFMDIKPRLHANCCPSFTGLCHVDKIIMKK 120
 ::: ::| : ::||| |::| |::| : ::| : ::|
Db 184 LGL-PYP-----GSPALEKLAVGSDPTRYALPRALLGRDCDFSFGLTAARAIAETL 236
 ::: ::| : ::||| |::| |::| : ::| : ::|
OY 121 EKESEIGEKQLLSAADIAATVQHTMACHLVKTRTHAILFCCKORDLLPONNAVTVASGV 180
 ::: ::| : ::||| |::| |::| : ::| : ::|
Db 237 TTDD-----ARRDLAAGVQAIALROLSERVDPAMKLTK--DSHDPEDLFVVAGV 285
 ::: ::| : ::||| |::| |::| : ::| : ::|
OY 181 ASNFYIRALBITLNATQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLGIHDEIGIRYE 240
 ::: ::| : ::||| |::| |::| : ::| : ::|
Db 286 AANGAVRAALADCEKNGSFAAPPLAYCTDNAAAMIALGAERL--ALGIFDDLDAIRA-R 342
 ::: ::| : ::||| |::| |::| : ::| : ::|
OY 241 PKCPL 245
 ::: ::| : ::||| |::| |::| : ::| : ::|
Db 343 PRWPL 347

RESULT 8

H83572
O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO1
C|Species: Pseudomonas aeruginosa
C|Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C|Accession: H83572
D.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Iaribig, K.; Lim,
-; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A|Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathoge
A|Reference number: AB2950; MUID:20437337; PMID:10984043
A|Accession: H83572
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-341 <STO>
A|Cross-references: GB:AE004494; GB:AE004091; NID:g9946446; PIDN:AAG03969.1; GSPDB:GN0011
C|Genetics:
A|Experimental source: strain PAO1
A|Gene: gcp; PA0580
C|Superfamily: O-sialoglycoprotein endopeptidase

Query Match 21.0%; Score 291.5; DB 2; Length 341;
Best Local Similarity 35.4%; Pred. No. 1, 5e-18;
Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;

[illegible]

RESULT 9
 G70369
 sialoglycoproteinase - Aquifex aeolicus
 C/Species: Aquifex aeolicus
 C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
 C/Accession: G70369
 R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
 Nature 392, 353-358, 1998
 A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A/Reference number: A70300; MUID:98196666; PMID:9537320
 A/Accession: G70369
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-335 (AOP>
 A/Cross-references: UNIPROT:O66986; GB:AE000708; NID:g2983356; PIDN:AAC06951.1; PID:g2929
 A/Experimental source: strain VFS
 C/Genetics:
 A/Gene: gcp
 C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match	21.0%;	Score 290.5;	DB 2;	Length 335;
Best Local Similarity	34.3%;	Pred. No. 1.8e-18;		
Matches	85;	Conservative 46;	Mismatches 82;	Indels 35;
			Gaps	10;
Qy	1	MEAHALTRLTNNKPEPFPLVLLISGGHCLTALVQVSDPFLIAGSLIDAPGMDLTKVAR	60	
Db	112	LEGGIYVFLPEKYEYFPFLAIIISGHTDLVYVDPGRYDPLGSLDDAVEAYDKAKM	171	
Qy	61	LSLTKHECSTMSGGKAIEMIAKQGNRFHFDIKPPLHAKKNCDFSFTGLQHTVKIIMKK	120	
Db	172	IGL-GYV-----GGPIIDRLAKGKGL-YPLPRLMEEGNLNFSFSLK---TALINTLL	220	
Qy	121	EKEEGIEKGQILSSADIIATVQHTMACHLVKTRTHAILFCQBDLLPONNAVLVASGV	180	
Db	221	KKEKNVAK-----EDIAVSFOETVAIILEKS---LWAKKGTGIMR-----LVVVGV	265	
Qy	181	ASNFYIRRAEILTNALQ---CTLLCPPRRLCTNGIMIAMNGIERLPAAGIILHIDEGI	237	
Db	266	SANSRLR---EVFKKAOEYGFELIYIHPSPISLTDNALMIAVAGMERFRRGVAVADLVNP	321	
Qy	238	RYEKCP	245	
Db	322	--QPNIP	327	

RESULT 10
H64074
O-salaloglycoprotein endopeptidase (EC 3.4.24.57) - Haemophilus influenzae (strain Rd KW)
N;Alternate names: salaloglycoproteinase
C;Species: Haemophilus influenzae

C|Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C|Accession: H64074
R|Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fife, L.D.; Fritchman, J.C.; Fuhrman, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A|Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter, A.
A|Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A|Reference number: A64000; MUID:95350630; PMID:7542800
A|Accession: H64074
A|Status: nucleic acid sequence not shown; translation not shown
A|Molecule type: DNA
A|Residues: 1-342 <TIGR>
A|Cross-references: UNIPROT:P43764; GB:U32735; GB:U42023; NID:g1573509; PIDN:AAC22187.1;
C|Superfamily: O-6-alkylglycoprotein endopeptidase
C|Keywords: hydrolase; metalloprotease

	Query Match	20.9%	Score 289.5;	D _B 2;	Length 342;
	Best Local Similarity	35.7%;	Pred. No. 2,3e-18;		
	Matches	82;	Conservative	30;	Mismatches 95; Indels 23; Gaps 7;
Qy	1	MEANAATRL-TNKVEFPPLVLLISGGHCLLAVGVSDPFLIGKSLDIPAGMDLVKAR	59		
		: : : :			
D _b	112	MEGHILAPMLTDNDSHPFVALVLVSGHTQLVRVDVGKKEVIGESITDDAAGEAFDKTKX	171		
		: : : :			
Qy	60	RLSLKHEPCSTMSGGKAIEHLAKOGNRPHFDIKPLIHAAKNDFSPFGQHVTDKIMK	119		
		: : : :			
D _b	172	LIGL-DYF-----GGAALSRLEKETPNRFTFRPMYTBAGADFSFGSKTTPAANYNQ	224		
		: : : :			
Qy	120	KEKEEG--TEKGQILSSADIATVAOHTMACHLVKKTHRAILECOKDILLPÖNNAVIVAS	177		
		: : : :			
D _b	225	AIKNEGELIEQ-----TKADIAYAFODAVVDTLA-----IKCK-RALKETGYKRVIYA	271		
		: : : :			
Qy	178	GVSASFPIYRRRLLETLTNAQTCTLCPRCLCPDNNGIMIAMNERBRAG	227		
		: : : :			
D _b	272	GGVSANKKLRTSLAHLMONTGSEVFYPOPOCFDNDGAMIAIYTGRLRKQG	321		
		: : : :			

RESULT 11
AI0079
probable glycoproteinase gcp [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C/Date: 02-Nov-2001 #sequence _revision 02-Nov-2001 #ext_change 09-Nov-2001
C/Accession: AI0079
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibbally, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, F
Native 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AI0079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <K0U>
A:Cross-references: GB:AL590842; PIDN:CAC89500.1; PID:gl5978736; GSPDB:GN00175
C:Genetics:
A:Gene: gcp
C:Superfamily: O-sialoglycoprotein endopeptidase

	Query Match	20.2%	Score 279.5;	DB 2;	Length 337;
	Best Local Similarity	32.8%	Pred. No. 1,8e-17;		
	Matches	81;	Conservative	35;	Mismatches 102; Indels 29; Gaps 7;
Oy	1	MEAHATRLR..TNKVEFPPLVLLISGGHCLLALVQGVSDPFLGLGSKSIDIPAGMDLVYAR	59		
Db	112	MEGHIALPMLEBNABEPFVALVLVSGHTQLLSVTNIGEVLLLGESVDADAGEAFDITAK	171		
Oy	60	RSLTIKHPECSTMSGKAIEHLAKOONRFAPDIKPELIHAKNCDSPFTYGQHVTDKIIMK	119		
Db	172	LGL-LDY-----GGPMISRMAQQGTVERFTTPRPMTDRPGIDFSGLKTFAANTIRA	224		
Oy	120	KEKEGIENKGQILSSADIATAVQHTMACHLVKRTIRAILFCOKORDLLPONN-AVLVASG	178		

Db 225 NGDDD-----QTRADIARAFEDAVVDTLAIKSKRA-----LDQTFKRLVAVG 267

Qy 179 GVASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIANWGIERLRAGLGIHIDIGIR 238

Db 268 GVSANQTRILRLKLAIDMQRKGGEVFYARPEFCTDNGMIAYAGVRLRSN---LNSLSVVS 324

Qy 239 YEPKCP 245

Db 325 VRPRWPL 331

RESULT 12

AG0892

probable glycoproteinase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S

C/Species: Salmonella enterica subsp. enterica serovar Typhi

A/Note: this species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: AG0892

R/Parikhil, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulé, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A/Reference number: AB0502; PMID:21534947; PMID:11677608

A/Accession: AG0892

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-337 <PAR>

A/Cross-references: GB:AL513382; P1DN:CAD07733.1; P1D:q16504285; GSPDB:GN00176

C/Genetics:

A/Gene: STY3387

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 20.2%; Score 279.5; DB 2; Length 337;

Best Local Similarity 32.9%; Pred. No. 1.8e-17;

Matches 81; Conservative 35; Mismatches 103; Indels 27; Gaps 8;

Qy 1 MEAHALITRL-TNKVEPFLVLLISGHCIALVGVSDPFLIGKSLDIAPGDMIDKVAR 59

Db 112 MEGHLAPMLBNDPPFPFVALLVSGHQLISVIGIYELGSLIDDAAGAPDKTAK 171

Qy 60 RLSTIKHPECSTWGGKAI EHLAKQGNRFHDIKPLHNAKNCDSFGTGLQHTDKITMK 119

Db 172 LIGL-DYP-----GGPLSKMAAGTGRFVPRPMTDRPGIDFSFGSKTFAANTIR 224

Qy 120 KEKEGIEKGQILSSADIAATVOHTMACHLVKTRTRALIFCKQRDLFPONNAVIVASGG 179

Db 225 NGDDE-----QTRADIARAFEDAVVDTL-----MICK-RALSTGPKRLVMAAG 268

Qy 180 VASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIANWGIERLRAGLGIHIDIGIR 239

Db 269 VSANRTLRALKLAEMQKRGGEVFYARPEFCTDNGMIAYAGVPRKA--GVTADL-GVTV 325

Qy 240 EPKCP 245

Db 326 RPRWPL 331

RESULT 13

C/Accession: C91122

probable O-sialoglycoprotein endopeptidase [imported] - Escherichia coli (strain O157:H7

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gaeswara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A/Reference number: A53625; PMID:21156231; PMID:1128796

C/Accession: C91122

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-337 <HAV>

A/Cross-references: GB:BA000007; P1DN:BAJ3730.1; P1D:q13363420; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain R1MD 0509952

C/Genetics:

A/Gene: EC83947

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 20.1%; Score 278.5; DB 2; Length 337;

Best Local Similarity 33.3%; Pred. No. 2.2e-17;

Matches 82; Conservative 36; Mismatches 101; Indels 27; Gaps 9;

Qy 1 MEAHALITRL-TNKVEPFLVLLISGHCIALVGVSDPFLIGKSLDIAPGDMIDKVAR 59

Db 112 MEGHLAPMLBNDPPFPFVALLVSGHQLISVIGIYELGSLIDDAAGAPDKTAK 171

Qy 60 RLSTIKHPECSTWGGKAI EHLAKQGNRFHDIKPLHNAKNCDSFGTGLQHTDKITMK 119

Db 172 LIGL-DYP-----GGPLSKMAAGTGRFVPRPMTDRPGIDFSFGSKTFAANTIR 222

Qy 120 KEKEGIEKGQILSSADIAATVOHTMACHLVKTRTRALIFCKQRDLFPONNAVIVASGG 179

Db 223 --RDNGTD-----QTRADIARAFEDAVVDTLMIKCKRAL-----DLTGFKR--LVWAGS 268

Qy 180 VASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIANWGIERLRAGLGIHIDIGIR 239

Db 269 VSANRTLRALKLAEMQKRGGEVFYARPEFCTDNGMIAYAGVPRKA--GATADL-GVSV 325

Qy 240 EPKCP 245

Db 326 RPRWPL 331

RESULT 14

B85967

probable O-sialoglycoprotein endopeptidase ysjd [imported] - Escherichia coli (strain O1:

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: B85967

R/Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, J.

iller, L.; Grobeck, E.J.; Davis, N.W.; Lin, A.; Dimantata, E.; Potamocists, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; PMID:21074935; PMID:11206551

A/Accession: B85967

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-337 <STO>

A/Cross-references: GB:AB005174; NID:q12517643; P1DN:AG58198.1; GSPDB:GN00145; UMG:2443

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Gene: ysjd

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 20.1%; Score 278.5; DB 2; Length 337;

Best Local Similarity 33.3%; Pred. No. 2.2e-17;

Matches 82; Conservative 36; Mismatches 101; Indels 27; Gaps 9;

Qy 1 MEAHALITRL-TNKVEPFLVLLISGHCIALVGVSDPFLIGKSLDIAPGDMIDKVAR 59

Db 112 MEGHLAPMLBNDPPFPFVALLVSGHQLISVIGIYELGSLIDDAAGAPDKTAK 171

Qy 60 RLSTIKHPECSTWGGKAI EHLAKQGNRFHDIKPLHNAKNCDSFGTGLQHTDKITMK 119

Db 172 LIGL-DYP-----GGPLSKMAAGTGRFVPRPMTDRPGIDFSFGSKTFAANTIR 222

Qy 120 KEKEGIEKGQILSSADIAATVOHTMACHLVKTRTRALIFCKQRDLFPONNAVIVASGG 179

Db 223 --RDNGTD-----QTRADIARAFEDAVVDTLMIKCKRAL-----DLTGFKR--LVWAGS 268

Qy 180 VASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIANWGIERLRAGLGIHIDIGIR 239

Db 269 VSANRTLRALKLAEMQKRGGEVFYARPEFCTDNGMIAYAGVPRKA--GATADL-GVSV 325

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_blue_p2n model

Run on: February 16, 2005, 13:19:07 : Search time 3172.37 Seconds
(without alignments)
3203.653 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALTRLNKVEPFLV.....DISKEVGBASIKVQLKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-Q/cgm2_1/USPTO.spool/US10649273/runat_14022005_114703_16411/app_query.fasta_1.1429
-DB=EST -QPM=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=humand40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NOR=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273.QCGN_1.1_6799_@runat_14022005_114703_16411 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOBERRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1263	91.2	870	5	BQ423651 AGENCOURT
2	1224	88.4	2284	3	AK045669 Mus muscu
3	1216	87.8	1622	3	AK011265 Mus muscu
4	1097	79.2	640	5	BQ636028 hdo3d11.Y
5	1075	77.6	852	5	BX391919 BX391919
6	1071	77.3	1605	3	BC030671 Mus muscu
7	1053	76.0	701	2	BE740611 BE740611
8	944	68.2	922	5	BQ961028 AGENCOURT
9	925	66.8	658	7	CF362328 CF362328

10	916	66.1	822	7	CF257246	CF257246 pha008 90
11	838	60.5	537	7	CK941819	CK941819 4065407 B
12	830	59.9	697	6	CB272391	CB272391 m517912
13	819.5	59.2	792	5	BU403563	BU403563 604138456
14	816	58.9	730	7	CN823245	CN823245 Oa_splbn
15	789	57.0	545	1	AV602901	AV602901 AV602901
16	786	56.8	866	5	BU127463	BU127463 603114407
17	784	56.6	490	6	CB852881	CB852881 UI-CF-FNO
18	749	54.1	735	7	CK365185	CK365185 AGENCOURT
19	749	54.1	1173	6	CD508917	CD508917 CNA93-E05
20	746	53.9	723	5	BU261251	BU261251 603502215
21	722	52.1	812	5	BU246489	BU246489 603784202
22	717	51.8	634	2	AM601179	AM601179 RC1-B7025
23	709	51.2	484	1	AJ670918	AJ670918 AJ670918
24	708	51.1	735	6	CA057753	CA057753 sea1rpb54
25	707.5	51.1	701	5	BU621780	BU621780 UI-H-FLI
26	704	50.8	749	7	CK982692	CK982692 4115331 B
27	698	50.4	696	5	BU302606	BU302606 603739448
28	692	50.0	579	7	CK819035	CK819035 1106405.Y
29	691.5	49.9	789	5	BU242187	BU242187 603781023
30	683	49.3	909	5	BX756548	BX756548 BX756548
31	682	49.2	548	7	CO880741	CO880741 Bvgen_09
32	671	48.4	682	2	BB043703	BB043703 BB043703
33	668	48.2	706	5	BU202465	BU202465 603949052
34	666	48.1	869	5	BX754527	BX754527 BX754527
35	647	46.7	919	7	CF407294	CF407294 CH3#047 E
36	646	46.6	878	5	BX776940	BX776940 BX776940
37	645	46.6	533	4	BM126453	BM126453 1106405.Y
38	638	46.1	1082	5	BX359023	BX359023 BX359023
39	625	45.1	506	2	BF415802	BF415802 UI-R-CA1
40	618.5	44.7	879	5	BU256052	BU256052 603776167
41	614	44.3	861	5	BU246158	BU246158 603779905
42	611	44.1	1171	5	BU261605	BU261605 603501763
43	604	43.6	424	1	AA273921	AA273921 V999C03.T
44	602	43.5	413	1	AA589724	AA589724 V178C10.T
45	600	43.3	863	5	BU376295	BU376295 603808890

ALIGNMENTS

RESULT 1
BQ423651
LOCUS
DEFINITION BQ423651 870 bp mRNA linear EST 23-MAY-2002
AGENCOURT_7790948 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065828
5', mRNA sequence.
ACCESSION BQ423651 GI:21118966
VERSION BQ423651.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 870)
NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC/DC/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: llam13342 row: 1 column: 21
High quality sequence stop: 710.
Location/Qualifiers
1..970
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6065828"

FEATURES

source
1..970
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6065828"

/tissue type="melanotic melanoma"
 /lab host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_72"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI,
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

ORIGIN

Alignment Scores:

Pred. No.:	1.9e-136	Length:	870
Score:	1263.00	Matches:	248
Best Similarity:	98.42%	Conservative:	1
Best Local Similarity:	98.02%	Mismatches:	2
Query Match:	91.19%	Indels:	2
DB:	5	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x BQ423651 (1-870)

```

QY      1 MetGluAlaHisAlaLeuThrIleArgLeuPheAsnValGluPheProPheLeuVal 20
DB      |||
QY      112 ATGAGAGGCTCAAGCACTTAATAATGAGTTGACCAATTAAGTAAGTAATTCCTTTTAACT 171
DB      |||
QY      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB      |||
QY      172 CTTTGTGATTTGTGAGGTCACTGTCTGTGGCATTAAGTTCAGAGAGCTTCAGATTTTCG 231
DB      |||
QY      41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB      |||
QY      232 CTTCTTGGAAAGTCTTTGACATAGCACAGGTGACATGCTTGAACAAGGTGCAAGAAGA 291
DB      |||
QY      61 LeuSerLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 80
DB      |||
QY      292 CTTTCTTTAAATAAATCAATGAGGTCTCCACATGAGGTGGGGAAGCCATTAAGAACT 351
DB      |||
QY      81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleIleIleIleIleIleIleIleIle 100
DB      |||
QY      352 TTGGCCAAACAGCAAGAAATGATTTTTCATTTGACATCAACCTCCCTTGCATATGCTAA 411
DB      |||
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleAspIleIleIleIleIleIle 120
DB      |||
QY      412 AATTGTGATTTTCTTTTACTGACACTTCAACAGCTTACTGATTAATATATGAAGAAAG 471
DB      |||
QY      121 GlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 140
DB      |||
QY      472 GAAAAAGAGAGAGATTTGAGAGAGGAGCAAAATCCGTCTTCAGCAGCAGCAATTCGTC 531
DB      |||
QY      141 ThrValGlnHisThrMetAlaCysHisIleuValIleAspIleIleIleIleIleIleIle 160
DB      |||
QY      532 ACAGTACAGCACAATATGCAATGCTTCTTGTGAAAAGAACATCGGGCTATTCGCTTT 591
DB      |||
QY      161 CysIleGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyIleVal 180
DB      |||
QY      592 TGTATGAGCAGAGAGCTTGTACCTCAAAATATATGCAATGCTGTTGATCTGTGTGTC 651
DB      |||
QY      181 AlAspSerAspPheIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 200
DB      |||
QY      652 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAACGCAACACAGTGCAC 711
DB      |||
QY      201 LeuLeuCysProProProArgLeuLeuCysThrAspAsnGlyIleMetIleIleIleIleIle 220
DB      |||
QY      712 TTGGTGTGTCTCTCTCCAGACTATGCACTGATTAATGAGCATTAATATGATGAGAAATG 771
DB      |||
QY      221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgIleGlu 240
DB      |||
QY      772 ATTGAAGAAGCTACGTCGTGGCTGGGCACTTTTACATGACATGAAGAGCATCGGCTATGA 831
DB      |||
QY      241 -ProIleCysPro-LeuGlyIleValAspIleSerIleIleIleIleIleIleIleIle 251
DB      |||
DB      832 ACCAAATNGTCCTCTTGGAGTAGACATATCAAAA 866
  
```

LOCUS

DEFINITION

AK045669 2284 bp mRNA linear HTC 03-APR-2004
 Mus musculus adult male corpora quadrigemina cDNA, RIKEN
 full-length enriched library, clone: B30219017 Product: similar to
 PUTATIVE STALOGLYCOPROTEINASE TYPE 2 [Homo sapiens], full insert
 sequence.

ACCESSION

VERSION

AK045669.1 GI:26337528

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLES

JOURNAL

MEDLINE

PUBMED

REFERENCES

AK045669 2284 bp mRNA linear HTC 03-APR-2004
 Mus musculus adult male corpora quadrigemina cDNA, RIKEN
 full-length enriched library, clone: B30219017 Product: similar to
 PUTATIVE STALOGLYCOPROTEINASE TYPE 2 [Homo sapiens], full insert
 sequence.

1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wachihi, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, K., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed capillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

4 The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

5 The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6 (bases 1 to 2284)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanakawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kurata, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct SubMISSION
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>
Location/Qualifiers

FEATURES

SOURCE

1. 2284
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM_DB:B230219017"
/db_xref="taxon:10090"
/clone="B230219017"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
221..1465
/note="unnamed protein product; putative
similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo
sapiens) (SPTK|Q9H4B0, evidence: FASTY, 80%ID, 100%length,
match=1242)"
/protein_start=1
/protein_id="BAC32450.1"
/db_xref="GI:26337529"
/translation="MIMLRRTGAI PKPKSKVYGLRFRSVHPTLSCHLVGIER
SCDTGAAYDETGVNCEALHSQTQVHLKKGIVPFAQQLHRENTIRIVETLSAC
RITPSDLISAIATTKIPGLALSIVGLSPSLQVNOFKPFIPIHMEHALITRLTNK
VEPPLVLLISGHCLLALVQGVSDFLVGLSLDIAKMDLKVARSLLKHPGST
MSGKALIEQLADGNRHFHTINPMDNKNCFSTQGHITDRLITKEPEGIEKG
OISSADIEAIVOHATACHLAKRTRAILLCKOKNLLSPNANVLYVSGVSNLYIR
KALEIVANATCTCLCPPRCTDNGIMIANNGIERLRAAGLGVLDHVEDIRPKPL
GVDSIRVABEAIVPRKLKML"

CDS

polysignal
polysite
/note="putative"
/note="putative"

ORIGIN

Alignment Scores:
Pred. No.: 2,756-131 Length: 2284
Score: 1224.00 Matches: 233
Percent Similarity: 92.51% Conservative: 14
Best Local Similarity: 87.27% Mismatches: 20
Query Match: 88.38% Indels: 0
DB: 3 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x AK045669 (1-2284)

QY 1 MetGluAlahisAlaLeuThrIleArgLeuThrAsnIlyValGluPheProPheLeuVal 20
DB 662 ATGAGAGCTCAGCAGCTACTATTAGGCTCACCAATAAGTGAATTTCTTTTATGTT 721
QY 21 LeuLeuIleSerGlyGlyIleHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 722 CTTTGTATTTCTGGCGGCTCCTGCTGTGGCATTAGTCCAAAGGCTGTTCCGATTTCTGTG 781
QY 41 LeuLeuGlyIlySerLeuAspIleAlaIleProGlyAspMetLeuAspIlyValAlaIleArg 60
DB 782 CTCTTTGGAGAGCTTTTGGACATAGCACCAGGCGACATGCTTGACAGAGTGGCAAGAGA 841
QY 61 LeuSerLeuIleIlySerProGlnCysSerThrMetSerGlyGlyIlyValAlaIleGluHis 80
DB 842 CTTTCTTAAATCAACAATCAGAAATGTTCTCAATGAGAGTGGGAAAAGCTATAGACAG 901
QY 81 LeuAlaIlySerGlnGlyAsnArgPheHisPheAspIleIlySerProTolLeuHisAlaIly 100
DB 902 TTGGCCAAAGACGAAATAGATTCATTTTCTACTCAATCCACCTATGACAGATCTAAG 961
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIlyValIleMetIlyIly 120
DB 962 AATTGCAATTTTCTTCAACGAGACTTCAACATATTACTGATAGCTAATTAACACACAG 1021

QY 121 GluIySGluGluGlyIleGluIySGlyGlnIleLeuSerSerAlaAlaIleAla 140
DB 1022 GAAAAAGAAAGAGGCTTAGAAGGGCAATCTCTCATCAGCTCAGACATGCTGCT 1081
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIlyAspGlyThrHisAlaIleLeuPhe 160
DB 1082 GCGGTACAGACATGCAACAGCTGCCACTTGGCAAAAGAACACATCGCGCTATTCTGTTT 1141
QY 161 CyllySGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 1142 TCCAGCAGCAAAAATTTGCTCTCTCCAGCTAACGCGAGTATTATTATTCATTCGAGGTGTT 1201
QY 181 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1202 GCAAGTAACTTGATCATCCGAAAGCATTTGGAAATGTGCGAAATCCAGCAGTCCAGC 1261
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly 220
DB 1262 TTGTTGTGTCCACTCTCAAGACATGTCACATGACATCATGATTCATTCATGAGATGGA 1321
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgIyGlu 240
DB 1322 ATTGAAGATTACGTGCTGCTGGCGCTTTTACATGATGTAAGACATCCGATATGAA 1381
QY 241 PolyAspProLeuGlyValAspIleSerIlyGlyValGlyAlaSerIleIlyVal 260
DB 1382 CCAAAATGCTCTTGGAGATGACATATCCAGAAAGATTGCGAAGCTCCGCAATAAAGTA 1441
QY 261 ProGlnLeuIlyMetGluIle 267
DB 1442 CCGCATTAATAATGCACTT 1462

RESULT 3
AK011265
LOCUS
DEFINITION
MUS MUSCULUS 10 DAYS EMBRYO WHOLE BODY cDNA, RIKEN FULL-LENGTH
ENRICHED LIBRARY, CLONE:2610001M19 PRODUCT:SIMILAR TO PUTATIVE
SIALOGLYCOPROTEASE TYPE 2 (HOMO SAPIENS), FULL INSERT SEQUENCE.
AK011265
VERSION
KEYWORDS
SOURCE
ORGANISM
MUS MUSCULUS (HOUSE MOUSE)
REFERENCE
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subcloning of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
REFERENCE
PUBMED
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Komoto, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, R., Mochizuki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

REFERENCE	TITLE	JOURNAL	AUTHORS
4	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.		Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
5	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 1622)
	Adachi,T., Aizawa,K., Akahira,S., Akimura,T., Ariai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furukoshi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirokita,T., Horii,F., Imotani,K., Ishii,Y., Itchih,M., Izawa,M., Kasekawa,T., Kato,H., Kawai,U., Kojima,Y., Komuro,H., Kouda,M., Koya,S., Kuribara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shidota,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyra,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Yuramatsu,M. and Hayashizaki,Y.		Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
	Please visit our web site (http://genome.gsc.riken.jp/) for further details.		
	cDNA library was prepared and sequenced in Mouse Genome Encyclopaedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAAGATCCACAGAGCTGTTCCTTTTCTTTTCTTCA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTTCAGATTAAATTAATCATCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.		
FEATURES	SOURCE	TITLE	JOURNAL
CDS			
	location/Qualifiers		
	1..1622		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6J"		
	/db_xref="PANTOM_DB:2610001M19"		
	/db_xref="taxon:10090"		
	/clone="2610001M19"		
	/tissue_type="whole body"		
	/clone_lib="RIKEN full-length enriched mouse cDNA library"		
	/dev_stage="10 days embryo"		
	207..1451		
	/note="unnamed protein product, putative similar to PUTATIVE SIALOGLYCOPROTEIN TYPE 2 (Homo sapiens) (SPTR Q94B0, evidence: FASTV, 80%ID, 100length, match=1242)"		
	/codon_start=1		
	/protein_id="BAB27506.1"		
	/db_xref="GI:12847276"		
	/translation="MLMRFRAGALIPRPKSKYTGFLRRSVHPRTLSCHKLVIQITSCDTPGAADVDETGNVLGEALSLGYQLTKGVGPVPVAQDLHENIQRVEETLSAQRTSIDSAIAITTKPGIALSIGVALSGSTLDLVNQPKFPIDIMNEHAHALTIITLTKRVSEFPLILIGSCHLALVGQVSDFDLGLSDIAPDKMDPIKHMAHLISLIKPEGSTWEPKALIEOLAKGNRPHFTNPDMONAKNDPSFTGLOHTDGLITHKEEGEEKEMOIISADADIAAVOHATHRLAKTHRAILFCCKNTLSPANNALVYSGYSASLTLYIRKAELIVANATCTCLCPERLCCTONGIMIANNGIERLRAGLGVLHDVEDIKYBRCPPLGVDSIREVAEAIVPRLLKML"		

polya_signal	1605..1610	/note="putative"
polya_site	1622	
ORIGIN	/note="putative"	
Alignment Scores:	1,44e-130	1622
Pred. No.:	1216.00	232
Score:	92.13%	213
Percent Similarity:	86.89%	Mismatches: 21
Best Local Similarity:	87.80%	Indels: 0
Query Match:	3	Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AK011265 (1-1622)		
QY	1 MetGluh1ani5a1aleuthr11learg1leuthr1asn1y5val1giupheProphleuVal	20
Db	648 ATGAGAGCTCAGCGACCTGACATTAATAGGTGTCACCAATAAAGTAAGAAATTTCTTTTATGTT	707
QY	21 LeuLeu11se8rg1ygl1yhi5CyeLeuLeu1aleuVal1ging1yVal1Ser1aspheLeu	40
Db	708 CTTTGTGATTTCTGGCGGTCACCTGGCTGTGGCATTAATCCAAAGGTTGTTCCGATTTCTGTG	767
QY	41 LeuLeu1y1y5se8r1leu1aap11ala1Progi1yasphe1leu1aap1yval1alarg1g	60
Db	768 CTCCTTGGAGAGTTTGTGACATAGCACCCAGCCACATGCTTGACACAGGTGGCAAGGA	827
QY	61 LeuSer1leu11ely5hi5Progi1y5se8rThrMetSer1ygl1y5ala11eq1uhi5	80
Db	828 CTTTCTTAAATCAACAATCCAGATGTTTCTTCAATGAAGTGGTGGAAAAGCTATAGAACAG	887
QY	81 Leu1al1y5g1ng1ya5nar1yph1e1h1e1aap11ely5ProProLeu1hi5hi5al1y5	100
Db	888 TTGGCCAAAGACGGAAATAGATTCATTTTACTATTCATCAATCCACCTATGACGAATGCTAAG	947
QY	101 Asn1y5a8rPh1e8erPh1eThr1y1leu1g1hi5val1Thr1aap1y11e11em1ely5	120
Db	948 AATGCGCATTTTCTTTCACCGGACCTTCACATATTAATCTGATTAAGCTAATTAACACACAAAG	1007
QY	121 Glu1y5g1ng1y11e1y1y5g1ygl11eLeuSer1a1a1aap11e1al1a	140
Db	1008 GAAAGAAAGAGGCAATTGAGAAAGAGCAATCTGTCACTAGCTGCAGACATTCCTGCT	1067
QY	141 Thr1y1g1hi5thr1Met1a1Cye1hi5leuVal1y5ar1ythr1hi5arg1a1a11e1euphe	160
Db	1068 GCGGTACGACATGCAACAGCGTGCACCTTCGAAAGAAACACATCGCGCTATTCTGTTT	1127
QY	161 Cys1y5g1nar1ga8r1leu1eupProgi1n8a8n1a1Val1leuVal1A1aSer1ygl1yVal	180
Db	1128 TGCAGACGAAAAATTTGCTCTCTCCAGCTAACCCAGATTAAGTTGTAATTCGAGAGTGT	1187
QY	181 AlaSer1a8rPh1y1r11earg1Arg1Ala1eug11uleu1Thr1asn1a1aThr1g1nc1yThr	200
Db	1188 GCAAGTAATCTGTACATCCGAAAGACATTGAAATTTGTCCGAATTCACACGCAATGCAAG	1247
QY	201 LeuLeu1y5ProPro1a8r1leu1Cye1Thr1asp1ng1y11em1e11eal1a1TP1a8ng1y	220
Db	1248 TTGTGTGTTCACCTCCCAAGCTGTGACATGACATGSCATCAATGATTGCATGGAATGGA	1307
QY	221 T1e1gl1u1arg1leu1a1g1y1leu1g1y11e1euh1a8p11e1gl1y11earg1y1glu	240
Db	1308 ATTGAAAGATTAACCTGCTGCTGGCGCTTTTATCATGATGTGAAGAGCATCCGATATGAA	1367
QY	241 Pro1y5Cys1ProLeu1yVal1a8p11eSer1y5g1yVal1g1y1u1a1Ser11ely5Val	260
Db	1368 CCAAAATGCTCTCTTGAGTACATATCCAGAAAGTTCGAGAAAGCTGCCATTAAGATA	1427
QY	261 Progi1n1euh1y5metGlu11e 267	
Db	1428 CCGCGATTAAATAATGCACTT 1448	

B0636028
LOCUS B0636028 640 bp mRNA linear EST 15-JUL-2002
DEFINITION hd03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
ACCESSION B0636028.1 GI:21760487
VERSION B0636028.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 640)
Wistow, G., Bernshtein, S.L., Wyatt, M.K., Ray, S., Behal, A.,
Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.
Expressed sequence tag analysis of human retina for the NEBank
Project: Retbindin, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts
Mol. Vis. 8 (4), 196-204 (2002)
JOURNAL MEDLINE
PUBMED 12107411
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 03 row: d column: 11
Seq primer: M13R1 reverse primer (ABI).
Location/Qualifiers
1..640
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hd03d11"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retina cDNA (Un-normalized,
unamplified): hd/he"
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
was dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was constructed at
Bioserve Biotechnology (Laurel MD) essentially following
the protocols of the SuperScript Plasmid System full
details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor
[5'-pGACTAGTTCTAGATCCGACGCGCCG(T)15-3']. EST analysis
was performed on the unamplified library at the NIH
Intramural Sequencing Center (NISC)."

ORIGIN
Alignment Scores:
Pred. NO.: 2.94e-117 Length: 640
Score: 1097.00 Matches: 209
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.21% Indels: 0
DB: 5 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x B0636028 (1-640)

QY 59 AArgatgleuSerleuIlelyshisProgluCyserThmetSerGlylysalalle 78
Db 3 AGAAGACTTCTTTATATAAATCATCCAGAGTCTCCACATAGAGTGGGAAGCATA 62
QY 79 GtuhstleuAlaIyGlnGlyasnArgPheHispheAspIlelyeProProleuHisHis 98
Db 63 GAACATTGGCCAAACAGAAATAGATTTCATTGACATCAAACTCCCTTGCAAT 122

QY 99 AAlaIyAsnCysAspPheSerPheThrGlyleuGlnHisValThrAspIleIleIle 118
Db 123 GCTAAATAATTCATTTCTTTTCTTTTACTGACCTCAACAGCTTACTGATAAATAATATG 182
QY 119 lYelYegIuYegIuGluGluGlyIleGluYegIyGlnIleleuSerSerAlaIaaspIle 138
Db 183 AAAAAGAAAAGAGAGAGATGATGAGAGGGGCAATCCTGCTTCAGACAGACAT 242
QY 139 AAlaIaThValGlnHisThrMetAlaCyHisIleuValIlysrgrHrHisAgaIalle 158
Db 243 GCTGCCACAGTACAGACACAAATGGATGATCTTGTAAGAACACATCGGCTAAT 302
QY 159 leuPheCysIyegIuArgAspIleuProGlnAsnAsnAlaValIleuValaIaSerGly 178
Db 303 CTTTGTGTAAGAGAGAGACTTTGTAACCTCAAAATATGCACTGCTGTCATCTGGT 352
QY 179 GtYAlaIaSerAsnPheTyrlleaArgAlaIeugIuIleleuThrAsnAlaThrgIn 198
Db 363 GGTGTGCAAGTAACTTATATATCCGACAGCTCTGGAATTTTAACAACGACACAG 422
QY 199 CyThrIleuLeuCyAspProProAlaArgIeucYThrAspAsnGlyIleMetIlealTrp 218
Db 423 TGCACCTTGTGTGTCTCTCTCCACAGCTATGACATGATATGCAATTGATTGATGATG 482
QY 219 AangIyIleGluYrleuYrleuYrleuYrleuYrleuYrleuYrleuYrleuYrleuYrleu 238
Db 483 AATGATTAAGAAAGACTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542
QY 239 TYRGluprolyeCyAspProleuGlyValaAspIleSerIyGluValGlyIuAlaSerIle 258
Db 543 TATGAACCAAAATATCTCTCTGAGATGACATATCAAAAGAGTTGAGAGAGCTTCATA 602
QY 259 lYsValProGlnleuIySmetGluIle 267
Db 603 AAAGTACCAAAATTAATAATGAGATG 629

RESULT 5
BX391919/c 852 bp mRNA linear EST 28-APR-2004
LOCUS BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DK001Y802 3-PRIME, mRNA sequence.
ACCESSION BX391919
VERSION BX391919.2 GI:46846154
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 852)
Li, W.B., Gruber, C., Jesse, J. and Polyes, D.
Full-length cDNA libraries and normalization.
Unpublished (2001)
On May 13, 2003 this sequence version replaced gi:30611736.
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seque@genoscope.cns.fr, web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
1240.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0BA1035E07_CS0317_1&c=1240.r

FEATURES
source Location/Qualifiers
1..852
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK001Y802"

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
 DB 665 ATGAGAGGTCAAGCAGTCACTATTAGGCTCACCAATTAAGTAGATTCTTTTATGTT 724
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
 DB 725 CTTTGGATTTCTGGCGGTCACTGCTGTGGCATATGTCAGAGGTGTTCGATTTCTCTG 784
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
 DB 785 CTCTTGGAGAACTTTTGGACATGACACAGCGCACTGCTTGAAGAAGTGGCAAGAAGA 844
 QY 61 LeuSerLeuIleLysHisAspProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
 DB 845 CTTTCTTTAATCAAAATCCAGATGTTCTTACATGAGTGGTGGAAAGCTTATAGACAG 904
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100
 DB 905 TTGGCCAAAGACGGAATATGATTCATTTTACTATCAATCCACTATGACAGAAATCTAAG 964
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
 DB 965 AATTGCCATTTTCTTTCACGCGCACTTCAACATATTACTGATMACCTATATACACACAG 1024
 QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaLys 140
 DB 1025 GAAAGAAAGAGGCAATTTGCAAGAGGCAATCTGTCATCGCTGCAGACATTCCTGCT 1084
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
 DB 1085 GCGGTACAGCATGCAACAGCGTGCACCTTGGAAAGAACACATCGGCTATTCTGTTT 1144
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyGlyVal 180
 DB 1145 TGCAGACAGAAATTTGCTCTCTCCAGCTAACGCAATATTAGTTGATCTGAGGTGT 1204
 QY 181 AlSerSerAspPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
 DB 1205 GCAAGTAACTTGTATCTCCGAAGCATTTGGAATTTGCGCAATATGCAACGAGTGCACG 1264
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGly 220
 DB 1265 TTGTTGTCTCCACTCCAGACTGTGACTGACATGAGCATCATATTTGCA----- 1315
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTrgIu 240
 DB 1315 ----- 1315
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyGluAlaSerIleLysVal 260
 DB 1316 ---TGAATGTCCTCTTGAAGTAGACATATCCAGAGAGTTGCAAGAGTGCCTATAAAGTA 1372
 QY 261 ProGlnLeuLysMetGluIle 267
 DB 1373 CCGGCAATTAATAATGCGCACTT 1393
 RESULT 7
 BE740611 701 bp mRNA linear EST 15-SRP-2000
 LOCUS 601595739P1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3949640 5',
 DEFINITION mRNA sequence.
 BE740611
 VERSION BE740611.1 GI:10154603
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 701)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: csapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: LNCM84 row: n column: 09
 High quality sequence stop: 701.
 Location/Qualifiers
 1..701
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3949640"
 /issue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 9"
 /note="Organ: ovary; Vector: pORF7; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.6e-112 Length: 701
 Score: 1053.00 Matches: 206
 Percent Similarity: 99.05% Conservative: 3
 Best Local Similarity: 97.63% Mismatches: 0
 Query Match: 76.03% Indels: 2
 DB: 2 Gaps: 0
 US-10-649-273-2_COPY_148_414 (1-267) x BE740611 (1-701)
 QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
 DB 71 ATGAGAGGTCAAGCAGTCACTATTAGGCTCACCAATTAAGTAGATTCTTTTATGTT 130
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
 DB 131 CTTTGGATTTCTGGCGGTCACTGCTGTGGCATATGTCAGAGGTGTTCGATTTCTCTG 190
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
 DB 191 CTCTTGGAGAACTTTTGGACATGACACAGCGTGCATGCTTGAAGAAGTGGCAAGAAGA 250
 QY 61 LeuSerLeuIleLysHisAspProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
 DB 251 CTTTCTTTAATCAAAATCCAGATGTTCTTACATGAGTGGTGGAAAGCTTATAGACAT 310
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100
 DB 311 TTGGCCAAAGACGGAATATGATTCATTTTACTATCAATCCACTATGACAGAAATCTAAG 370
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
 DB 371 AATTGCCATTTTCTTTCACGCGCACTTCAACATATTACTGATMACCTATATACACACAG 430
 QY 120 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaLys 140
 DB 431 GGAACAAAGAGAGATTTGAGAGGCGCAATCTGCTTCCAGCAGACGATTTGCTG 490
 QY 140 IarHValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuP 160
 DB 491 CCACAGTACAGACACATGCGATGTCATCTTGTGAAAAAGACACATCGGCTATTCTGT 550
 QY 160 heCysLysGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyGlyVal 180

Db 551 TTGTAAAGCAGAGAGACTTGTTACTCAAAATATAGCAGTACTGGTTCATCTGGTGG 610
Qy 180 a1a1aSeraphenPhetTlLeaArgAlaLeuGlnIleLeuThraSnaAlaThrGlnCyt 200
Db 611 TCGCAAGTAACTCTTAATATCCGACAGAGCTCTGGAATTTTAAACAACGACACAGTGA 670
Qy 200 hrLeuLeuCySproProProArgLeuCyS 209
Db 671 CTTTGTGTGTCTCTCTCCAGACTATGCG 699

RESULT 8
LOCUS B0961028
DEFINITION AGENCOURT 8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423902
5' mRNA sequence.
ACCESSION B0961028
VERSION B0961028.1 GI:22376506
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 922)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP/Gazdar
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHC2603 row: d column: 15
High quality sequence stop: 584.
Location/Qualifiers
1..922
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6423902"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_18"
/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by King Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 3.65e-99 Length: 922
Score: 944.00 Matches: 207
Percent Similarity: 89.45% Conservative: 5
Best Local Similarity: 87.34% Mismatches: 16
Query Match: 68.16% Indels: 10
DB: 5 Gaps: 3
US-10-649-273-2_COPY_148_414 (1-267) x B0961028 (1-922)

Qy 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuAl 20
Db 207 ATGAGGCTCATGCACTACTAATTAGGTGACCAATAAGTAAGTAATTCCTTTTATGTT 266
Qy 21 LeuLeuIleSerGlnGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 267 CTTTGAATTTCTGAGAGTCACTGCTCTGTTGGCAATTAGTTCAAGAGCTTCAAGATTTCTG 326

Qy 41 LeuLeuGlnIleYsSerLeuAspIleAlaProGlnYAspMetLeuAspIleValAlaArgArg 60
Db 327 CTTCTTGGAAAGCTCTTTGACATAGCACAGGATGACATCTTGACAAAGGTGGCAAGAA 386
Qy 61 LeuSerLeuIleIleHisIleProGlnCyS-SerThreSerGlnGlyIleValAlaIleGlnI 80
Db 387 CTTTCTTTAATTAACATCCAGTCAGTGCCATCCATGATGGTGGGAAAGCCATAGACA 446
Qy 80 bleuAlaIysGlnGlnIleAsnArgPheHisIleAspIleIleYsProPheLeuHisIleAlaI 100
Db 447 TTTGGCCAAACAAGAAATAGATTTCATTTTACATCAAACTCCCTTCATCATCTTA 506
Qy 100 AsnCySAspPheSerPheThrGlnIleGlnIleHisValThrAspIleIleIleMetIleYs 120
Db 507 AATATTGATTTTCTCTTTACTGAGACTTCACACGTTACTGATTAATATATATGAAAA 566
Qy 120 sGlnIysGlnIleGlnIleGlnIleGlnIleLeuSerSerAlaIleAspIleAlaI 140
Db 567 GGAATAAAGAGAAAGTATTGAGAAAGGGAATCCGTCTTCAGACGACAGACTTCTCG 626
Qy 140 aThrValGlnHisThrMetAlaCysHisIleuValIleYsArgThrHisArgAlaIleLeuPh 160
Db 627 CACAGTACAGCACACAAATGCAATGTCATCTTGGAAGAAA-CATCGGGCTAATCTGTT 685
Qy 160 eCysIysGlnIleArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSer-GlyGlyI 180
Db 666 TTGTAAAGCAGAGAGACTTGTAACCTCAAAATATATGACAGTACGTGGTGATCTTGGGGGG 745
Qy 180 a1a1aSeraphenPhetTlLeaArgAlaLeuGlnIleLeuThr-AsnAlaThrGlnCyS 199
Db 746 TCGCAAGTAACTCTTAATATCCGCAACTGTCGAAATTTTCAAAAACCAACACGGGG 805
Qy 200 Thr-LeuLeuCySproProProArgLeuCySThr-AspAsnGlnIleMetIleAlaTrp- 218
Db 806 ACCTTGTGTGGGTCCCGCCCTCCCAACTATGCACTGGAATATGGG---CATTTATGATGC 862
Qy 219 -----AsnGlyIleGlnArg--LeuArgAlaGlyLeuGly 229
Db 863 TGGGGAAGGAATTGAAAAAATAACCTGCTGCGCTTGGGG 903

RESULT 9
LOCUS CF362328/c 658 bp mRNA linear EST 25-AUG-2003
DEFINITION 829596 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.
ACCESSION CF362328
VERSION CF362328.1 GI:34161882
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE 1 (bases 1 to 658)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Valliet,J.L., Fox,J.,
Wise,T.A., Noneman,D.J., Wray,J.E. and Keele,J.W.
TITLE A second set of porcine ESTs from a pooled-tissue normalized
library
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smtth@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: SK8015 row: H column: 12
Seq primer: TAGAAGGCAAGTCGAGG.
Location/Qualifiers
1..658
/organism="Sus scrofa"
/mol_type="mRNA"

/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="MARC 3P1G"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN

Alignment Scores:
Pred. No.: 3,71e-97 Length: 658
Score: 925.00 Matches: 174
Percent Similarity: 93.97% Conservative: 13
Best Local Similarity: 87.44% Mismatches: 12
Query Match: 66.79% Indels: 0
DB: 7 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x CF362328 (1-658)

QY 20 ValLeuLeuLeuSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39
DB 657 GTTCTTTGATATCTGGCGGTCATTGCTTTGGCATTGAGAGAGTTTCAGATTCTT 598
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
DB 597 CTGCTTCTGGACACTCTTTGGACATACACAGGTGACATCTTGGACAGGTAGCACA 538
QY 60 ArgLeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyLysAlaIleGlu 79
DB 537 AGACTTCTTAAATAAACATCCAGAGTCTCCACATGAGTGGGGAGGCCATTAGAA 478
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIle 99
DB 477 CATTGTCACAAACAGGAAATAGTGAATTTGATTCAACCTCCATCCATGCAAGTGA 418
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
DB 417 AAAATATGATGATTTCTTTCTTGACACTCAACAGTATTGATTAAGACAAATAGCAG 358
QY 120 LysGlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
DB 357 AAGGAAAAAGAGAGATGATGAGAGAGGCAATCTGCTTCAGCTGCAGACATTGCT 298
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
DB 297 GCTCAGATACAGCACACAGTACCTGCCTCATTTGCAAAAGAACATCATGCTATTCTG 238
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleLeuValAlaSerGly 179
DB 237 TTTTGCAAAACGAGAGACTTATATGTAAGTAATGAGATATGCTGATCTGAGGCT 178
QY 180 ValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
DB 177 GTTGCAAGTAACTTATATATAGAAAGCTTTAGAGTGTGACAAATGCAACAAATGC 118
QY 200 ThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218
DB 117 ACTCTGTTGTCTCTCTCCAGACTATGCACTGATTAATGCGATTATGATTCATGCG 61

RESULT 10
CF257246
LOCUS CF257246 822 bp mRNA linear EST 07-AUG-2003
DEFINITION phao08 g02 PHA-activated splenocytes Gallus gallus cDNA, mRNA
SEQUENCE.
CF257246
ACCESSION CF257246.1 GI:33490501
VERSION 1
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 822)
AUTHORS Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whitaker,C.J.,
Chausse,A.M., and Zoorob,R.
TITLE A collection of chicken ESTs from activated immune cells
JOURNAL Unpublished (2003)
COMMENT Contact: Zoorob R
UPR 1983
CNRS

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France
Tel: 33 1 49 58 35 00
Fax: 33 1 49 58 33 81
Email: zoorob@vjf.cnrs.fr.

FEATURES

source 1..822
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/cell_type="Splenocytes"
/clone_lib="PHA-activated splenocytes"
/note="Vector: pTriplex2"

ORIGIN

Alignment Scores:
Pred. No.: 5,75e-96 Length: 822
Score: 916.00 Matches: 175
Percent Similarity: 81.37% Conservative: 39
Best Local Similarity: 66.54% Mismatches: 49
Query Match: 66.14% Indels: 1
DB: 7 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x CF257246 (1-822)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 20
DB 35 ATGGAGGCTCACAGCACTTACCATCAGATGACAGACAGAAATGAAATTTCCCTTACTT 94
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 95 CTTTACTCTCCGAGAGTACTGCACTTGGCAGTGCAGAGAGACTTTCAGATTCTCTT 154
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 155 CTGCTTGACATCCATGATATACAGACAGAGTGAATCTTGGATTAAGTATAGCAAAAG 214
QY 61 LeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyLysAlaIleGluHis 80
DB 215 CTCTCTTATAGTAACACCCGAGATGCCACATGATGCCGCGGAGCAATGATAGCAGC 274
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLys 100
DB 275 CTGGCTCAAAACGAGAGACTGGCAACATCTTCAGCTTCCATCCATGCAAGTATGCT 334
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 335 AACTGTATTTCTTTCTCCGACTTCAGACCTTGTCAACAAAGCCATTCTTCGAA 394
QY 121 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 395 GAAAAAGAAAGAGATTAACAAGAGGAAATCTCTCCGCGTTAAGACATGCTGCT 454
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 455 GCTGACAGCAGTGTGCTGCTCTATATATTCACGAGACACACGAGCCATGCTCTTC 514
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleLeuValAlaSerGlyVal 180
DB 515 TGCATGAAAAACAGCATATTTATACCAAAATGCACTCTGTTATACAGAGAGATT 574
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 575 GCAATTAATCACTATATATGAAAGAGACTGCAAGACTCTGCAAAATGCAAAAGTTTGTCT 634
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220

Db 635 TTCTGCTCTCTCCAGAGCTGTGACCCGATTAATGATTAATGATGCAATGGC 694
Qy 221 TTTCTGCTCTCTCCAGAGCTGTGACCCGATTAATGATTAATGATGCAATGGC 694
Db 695 ATTAAAGAGTTCGTCGACGATGTCGATTTTATCACTAATGATGATCGCTACGAA 754
Qy 241 ProlycysProleuGlyValAspIleSerlysgIuValGlyAlaSerIleVal 260
Db 755 CCAAAAGCTCCCTTGATGATTAATTTCCAAAGATTGAA-GAGGATTCATCAAGTG 813
Qy 261 ProGlnLeu 263
Db 814 CCAAGACTA 822
RESULT 11
CK941819/c 637 bp mRNA linear EST 15-MAR-2004
LOCUS 4065407 BARC 10BOV Bos taurus cDNA clone 10BOV12_F24 3', mRNA
DEFINITION
CK941819
VERSION CK941819.1 GI:45456199
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 637)
Sontegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay,
G.P., Bosaik, S., Rubenfield, M. and Gasparre, L.C.
Production of EST from cDNA libraries derived from immunologically
activated bovine gut
Unpublished (2004)
CONTACT: Tad S. Sontegard
Bovine Functional Genomics Laboratory
Animal and Natural Resources Institute
Bldg. 200 RMA BARC-East, Beltsville, MD 20705, USA
Tel: 3015048416
Fax: 3015048414
Email: tads@anri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim alt -c -rim fasta. Vector identified
by cross match using options -mismatch 12 -mincore 18
Plate: 12 Row: F Column: 24
Seq primer: AGCGATACAAATTCACACAGG
High quality sequence stop: 637.
Location/Qualifiers
1..637
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="10BOV12_F24"
/sex="Male"
/issue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_id="BARC 10BOV"
/note="Organ: Small Intestine; Vector: pagen-1; Site 1:
EcoRV; Site 2: NotI; Equimolar amounts of mRNA extracted
from proximal jejunums of 18 and 21 wk old steers, and
distal ileums of 14 day old calves. Proximal jejunum
exposed to C. oncophora for 3 and 6 weeks, and distal
ileum exposed to C. parvum for 7 days"

ORIGIN

Alignment Scores:

Pred. No.: 5.09e-87 Length: 637
Score: 838.00 Matches: 153
Percent Similarity: 92.47% Conservative: 19
Best Local Similarity: 82.26% Mismatches: 14
Query Match: 60.51% Indels: 0

DB: 7 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x CK941819 (1-637)
Qy 82 AATySGInGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLysAsn 101
Db 632 GCCAAACAGGGAATAAGATTGATTTTCAGACCTCCCATGACAGCGCTAAAAAT 573
Qy 102 CysAspPheSerPheThnGlyLeuGlnHisValThrAspLysIleIleMetLysVal 121
Db 572 TGTGATTTTCTTTTCTTGAGCTCAACAGTTATTTGAATGAATGAATGAATGAATGA 513
Qy 122 LysGInGlyGlyIleGlyLysGlyIleLeuSerSerAlaAlaAspIleAlaThr 141
Db 512 AAAGAGAGAGATTCAGACAGGCGAGCTCTCTCTTCACTGACAGCATGCTGTGGC 453
Qy 142 ValGlnHisPheMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCys 161
Db 452 GTGACAGCACACCGTGCGCTGCACATTCGCAAAAGAACACATCGCTCTTGTCTGC 393
Qy 162 LysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAla 181
Db 392 AAGCAGAGAGGCTTCTTACGTACAGATACGAGTACGTGTATCTGTAGAGCGCTGCA 333
Qy 182 SerAsnPheTrpIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeu 201
Db 332 AGTAACTTATATATTCGAAAGACCTTGGAATTTGTACCAATGCACACAGTGCATTTG 273
Qy 202 LeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIle 221
Db 272 CTGTGCGCGCCGCCAGACTTGTGACAGCGGCTTATATATATATATATATATATATAT 213
Qy 222 GlnArgLeuArgAlaGlyLeuGlnIleLeuHisAspIleGlnGlyIleArgTrpGluPro 241
Db 212 GAAAGACTACCGTGTGCTGTGGCATTTTACCAACACAGAGAGGCAATCCGCTACGAA 153
Qy 242 LysCysProLeuGlyValAspIleSerlysgIuValGlyAlaSerIleValPro 261
Db 152 AATGTCTCTTGATGATTAATTCATCAAGAGATTGGAGAACTCTTAAAGTCCCA 93
Qy 262 GlnLeuLysMetGluIle 267
Db 92 AGATTAAATAATGAGATT 75
RESULT 12
CB272391 597 bp mRNA linear EST 24-FEB-2003
LOCUS ma157912.y1 McCarrey Eddy spermatoocytes Mus musculus cDNA clone
DEFINITION IMAGE:6445750 5' similar to TR:Q9VWD6 Q9VWD6 CG14231 PROTEIN. ;,
mRNA sequence.
CB272391 GI:28462714
VERSION CB272391.1
KEYWORDS Mus musculus (house mouse)
SOURCE EST.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 597)
McCarrey, J., Eddy, M., Marra, M., Hillier, L., Clifton, S., Pape, D.,
Martin, J., Wyllie, T., Dante, M., Bowers, Y., Theising, B., Gibbons, M.,
Ritter, E., Tsagarelis, R., Ronko, I., Maguire, L., Kennedy, S.,
Bennett, J., Waterston, R. and Wilson, R.
NIHES Mouse
Unpublished (2002)
TITLE JOURNAL
COMMENT CONTACT: McCarrey/Eddy NIHES Mouse
NIHES Mouse
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed and donated by J. McCarrey, Ph.D. (Southwest
Foundation for Biomedical Research, Dept. of Genetics) - excision

done by E.M. Eddy, Ph.D. (National Institutes of Health, National
Institute of Environmental Health Sciences).

MG1:2069710

Seq primer: Primer name ambiguous
High quality sequence stop: 419.

FEATURES

Source

```

1. 597
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:6445750"
/sex="male"
/tissue_type="spermatocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/clone_lib="McCarrey Eddy spermatocytes"
/note="Organ: testis; Vector: pBluescript SK+
(Stratagene); Site 1: XhoI; Site 2: EcoRI; cDNA oligo
dT-primed (5'-(GA)10-ACTAGTCTCAGTTTCTTTT-3') and
directionally cloned using 5' linkers 5'-AATCGCAGAG-3'
and 5'-CTCGTCCG-3'. Size selection of >400bp material
gives average insert size ranging from 1-2 kb. Library was
mass excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and retransformed
into DH10B. Library contains 96% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."

```

ORIGIN

Alignment Scores:

```

Pred. No.: 3.98e-86 Length: 597
Score: 830.00 Matches: 162
Percent Similarity: 91.80% Conservative: 6
Best Local Similarity: 88.52% Mismatches: 15
Query Match: 59.93% Indels: 0
DB: Gaps: 0

```

US-10-649-273-2_COPY_148_414 (1-267) x CB272391 (1-597)

```

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 20
DB 49 ATGAGAGCTACGACGACTGATTAAGGCTCACCAATAGTAGAATTTCTTTTAACTT 108
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerPheLeu 40
DB 109 CTTTGAATTTCTGGCGGTCACTGCTGTGGCATTAAGCAAGGTTTCCGATTTCTCG 168
QY 41 LeuLeuGlyIysSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB 169 CTCCTTGGAGAGTTTGTGACATAGCACGACGACATGCTTGACAAAGTGGCAAGAGA 228
QY 61 LeuSerLeuIleIysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 229 CTTTCTTAATCAAACTCCAGATGTTCTACATAGAGTGTGAAAAGCTATAGAACAG 288
QY 81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleIysProPheLeuHisAlaIys 100
DB 289 TTGGCCAAAGACGGAATAGATTCCATTTTACTATCAATCCATATGCAAGATGCTAAG 348
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetIysIys 120
DB 349 AATTGCAATTTTCTTTCACGGGACCTTCAACATATTAAGTAAGCTAATATACACAAAG 408
QY 121 GluIysGlnGlyIleGluIysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIys 140
DB 409 GAAAGAGAGAGAGCATTTGAGAGAGGCGCAATCTGTCAATGCTGACAGCATTTGCTCT 468

```

```

QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPhe 160
DB 469 GCGGTACACATGCCAACACACGCTGCCACTTGGCAAAAGAACATGCGCTATTCTGTTT 528
QY 161 CysIysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyIysVal 180
DB 529 TCCAAGCAGAAAATAATGCTCTCTCCAGCTAACGCAAGTATGATTATCTGAGAGTGT 588
QY 181 AlaSerAsn 183
DB 589 GCAAGTAAC 597

```

RESULT 13

BU0403563

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

1 (bases 1 to 792)
Boardman, P.E., Sanz-Bzquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
2235534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1..792
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, HiseX"
/db_xref="taxon:9031"
/clone="CHEST967014"
/dev_stage="16"
/lab_host="DH10B"
/clone_lib="CHEST967014"
/note="Organ: Limbs; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was bluntended, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996) 791, except that a significantly longer
reannealing hybridization was used."

FEATURES

source

ORIGIN

Alignment Scores:

```

Pred. No.: 1.01e-84 Length: 792
Score: 313.50 Matches: 164
Percent Similarity: 78.99% Conservative: 39
Best Local Similarity: 63.81% Mismatches: 53
Query Match: 59.17% Indels: 3
DB: Gaps: 1

```


ORGANISM

Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 545)
Takauega,A., Hirotsune,S., Itoh,R., Jitchono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
11713328

JOURNAL

CONTACT: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp

FEATURES

Location/Qualifiers
1..545
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="ELK1013A07"
/issue_type="kidney"
/dev_stage="fetus"
/lab_host="DHI08"
/clone_lib="Bos taurus kidney fetus"
/note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"

ORIGIN

Alignment Scores:

Pred. No.: 2.14e-81 Length: 545
Score: 789.00 Matches: 147
Percent Similarity: 90.11% Conservative: 17
Best Local Similarity: 80.77% Mismatches: 17
Query Match: 56.97% Indels: 1
DB: 1 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x AV602901 (1-545)

QY 85 G|YAsnAArgPheHisPheAspIlelySProProLeuHis-AtAlaYsAsnCyAspPh 104
DB 544 GGAATATGATTCATTTTATTCATTCANCTCCATGCAACGGTCTAATAAATTGTGATTT 485
QY 104 eSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluG 124
DB 484 TTTCTTTTCGCACTTCAACAGTTATGATTAAGATGATTAATGCAAAAGAAAANAGGA 425
QY 124 uGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnHis 144
DB 424 AGGTATGAGAGGAGGCGCTCTCTCTTCAGCTCGGACATTCGCTTGCAGTCCAGCA 365
QY 144 sThrMetAlaCyHisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGlnAr 164
DB 364 CACGGTGCCCTGCACATGCAAAAGAACACATGCTCTCTCTCTGTCGACAGCAGAG 305
QY 164 gAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPh 184
DB 304 AGGCTTCTTACATCAGAGTAAGCACTACTGTTGTATCTGAGGCGTCGCAAGTAATT 245
QY 184 eTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCyAsp 204
DB 244 ATATATCCGAAAGCCCTCGAAATGTGTACCAATGCAACAGTGCATTTGCTGTCTCC 185
QY 204 oProProArgLeuLeuCyThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArg 224
DB 184 GCCCCCAAGACTGCACTGCAACAGGCGTAAATGATTCATGAGAAATGTTGAAAGACT 125

QY 224 uArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrgIuProLysCyAsp 244
DB 124 AGCTGCTGAGCTTGCGCATTTTACACACACAGAGGCACTCCGTACGAAACCAAAATGTC 65
QY 244 oLeuGlyValAspIleSerLysGlyValGlyGlnAlaSerIleLysValProGlnLeu 264
DB 64 TCTTGAGTAAATATATATCAAAAGAAAGTGGAGAGCTGCTATAAAAGTCCCAAGATTAAA 5
QY 264 sMet 265
DB 4 AATG 1

Search completed: February 16, 2005, 21:04:47
JOD time : 3184.37 secs

Query Match 100.0%; Score 1385; DB 2; Length 414;
 Best Local Similarity 100.0%; Pred. No. 2.9e-110;
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAAHATIRLTNTKVEPFPVLVLLISGGHCLALVQVSDPFLLGKSLDIAFGMDLKVARR 60
 DB 148 MEAAHATIRLTNTKVEPFPVLVLLISGGHCLALVQVSDPFLLGKSLDIAFGMDLKVARR 207
 QY 61 LSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLAHAKNCDSEFTGLQHTYTKIIMKK 120
 DB 208 LSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLAHAKNCDSEFTGLQHTYTKIIMKK 267
 QY 121 EKEGIRKQGLSSADIAATVQHTMACHIVKTRATILFCQKORLLPONNAVVASGCV 180
 DB 268 EKEGIRKQGLSSADIAATVQHTMACHIVKTRATILFCQKORLLPONNAVVASGCV 327
 QY 181 ASNFYIRRAEILTNATQCTLLCPEPRLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
 DB 328 ASNFYIRRAEILTNATQCTLLCPEPRLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 387
 QY 241 PKCPGVDISKVEGASIKVPOLKMEI 267
 DB 388 PKCPGVDISKVEGASIKVPOLKMEI 414

RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.
 ID Q9H4B0
 AC Q9H4B0
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Putative sialoglycoprotein type 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Chen J.M., Fortunato M., Barrett A.J.;
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AJ295148; CACI4666.1; --
 DR MEROPS; M22.004; --
 DR Genew; HNCN:23075; OSGBPL1.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR00905; Peptidase M22.
 DR InterPro; IPR009180; Pept M22_Osialgl1.
 DR Pfam; PF00814; Peptidase M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase M22; 1.
 DR TRFRFams; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 439 AA; 48040 MW; 44849372C784B41F CRC64;

Query Match 98.1%; Score 1358.5; DB 2; Length 439;
 Best Local Similarity 91.1%; Pred. No. 5.8e-108;
 Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
 QY 1 MEAAHATIRLTNTKVEPFPVLVLLISGGHCLALVQVSDPFLLGKSLDIAFGMDLKVARR 60
 DB 148 MEAAHATIRLTNTKVEPFPVLVLLISGGHCLALVQVSDPFLLGKSLDIAFGMDLKVARR 207
 QY 61 LSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLAHAKNCDSEFTGLQHTYTKIIMKK 120
 DB 208 LSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLAHAKNCDSEFTGLQHTYTKIIMKK 267
 QY 121 EKEGIRKQGLSSADIAATVQHTMACHIVKTRATILFCQKORLLPONNAVVASGCV 180
 DB 268 EKEGIRKQGLSSADIAATVQHTMACHIVKTRATILFCQKORLLPONNAVVASGCV 327

DB 268 EKEGIRKQGLSSADIAATVQHTMACHIVKTRATILFCQKORLLPONNAVVASGCV 180
 QY 156 RAILFCCKORLLPONNAVVASGCVASNFYIRRAEILTNATQCTLLCPEPRLCTDNGIM 215
 DB 328 RAILFCCKORLLPONNAVVASGCVASNFYIRRAEILTNATQCTLLCPEPRLCTDNGIM 387
 QY 216 IAMNGIERLRAGLILHIDIGIRYRPPKCPGVDISKVEGASIKVPOLKMEI 267
 DB 388 IAMNGIERLRAGLILHIDIGIRYRPPKCPGVDISKVEGASIKVPOLKMEI 439

RESULT 3

Q6PB4 PRELIMINARY; PRT; 414 AA.
 ID Q6PB4
 AC Q6PB4
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold R.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaller G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hultky S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RA Strausberg R.;
 RL Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.
 RL EMBL; BC058172; AAH58172.1; --
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009180; Pept M22_Osialgl1.
 DR InterPro; IPR00905; Peptidase M22.
 DR Pfam; PF00814; Peptidase M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase M22; 1.
 DR TRFRFams; TIGR00329; gcp; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBACB CRC64;

Query Match 88.9%; Score 1231; DB 2; Length 414;
 Best Local Similarity 87.3%; Pred. No. 4.6e-97;
 Matches 233; Conservative 15; Mismatches 19; Indels 0; Gaps 0;
 QY 1 MEAAHATIRLTNTKVEPFPVLVLLISGGHCLALVQVSDPFLLGKSLDIAFGMDLKVARR 60
 DB 148 MEAAHATIRLTNTKVEPFPVLVLLISGGHCLALVQVSDPFLLGKSLDIAFGMDLKVARR 207
 QY 61 LSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLAHAKNCDSEFTGLQHTYTKIIMKK 120
 DB 208 LSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLAHAKNCDSEFTGLQHTYTKIIMKK 267
 QY 121 EKEGIRKQGLSSADIAATVQHTMACHIVKTRATILFCQKORLLPONNAVVASGCV 180
 DB 268 EKEGIRKQGLSSADIAATVQHTMACHIVKTRATILFCQKORLLPONNAVVASGCV 327

148 MEAHLATRLTNKVEPFPVLLISGHCALAVOGVSDPFLIGSKSLDIPAGMDLVKVAR 207
 61 LSLIKHPECSTMSGKAI EHLAKOGNRFHDIKPELHAKNCDPFTGLOHTDKIMKK 120
 208 LSLIKHPECSTMSGKAI EHLAKOGNRFHDIKPELHAKNCDPFTGLOHTDKIMKK 267
 121 EKEGIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKORDLLPQNNAVLVASGV 180
 268 EKEGIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKORDLLPQNNAVLVASGV 327
 181 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 240
 328 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 387
 241 PKCPGLGVDSKEVGEASIKVPOLKMEI 267
 388 PKCPGLGVDSKEVGEASIKVPOLKMEI 414

RESULT 4
 08BLB6 PRELIMINARY; PRT; 414 AA.

08BLB6; PRELIMINARY; PRT; 414 AA.
 01-MAR-2003 (TREMBLrel. 23, Created)
 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone: B20219017 product:similar to PUTATIVE STALOGLYCOPROTEASE TYPE 2.
 Name=Osegep11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI TaxID=10090;
 [1]
 [2]
 [3]
 [4]
 [5]

SEQUENCE FROM N.A.
 STRAIN=CS7BL/6J; TISSUE=Corpora quadrigemina;
 MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RIKEN FANTOM Consortium;
 "Functional annotation of a full-length mouse cDNA collection.";
 Nature 409:685-690(2001).
 [1]
 [2]
 [3]
 [4]
 [5]

SEQUENCE FROM N.A.
 STRAIN=CS7BL/6J; TISSUE=Corpora quadrigemina;
 MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.15100;
 Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Komno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";
 Genome Res. 10:1617-1630(2000).
 [1]
 [2]
 [3]
 [4]
 [5]

SEQUENCE FROM N.A.
 STRAIN=CS7BL/6J; TISSUE=Corpora quadrigemina;
 MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 Saitoh R., Itoh M., Aizawa S., Sasaki N., Carninci P., Komno H., Akiyama J., Nishi K., Kitagawa T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto K., Matsumoto H., Sakauchi S., Ikegami T., Kashiwagi K.,

Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Okazaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsuzura S., Kawai J.,
 "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";
 Genome Res. 10:1757-1771(2000).
 [6]
 [7]
 [8]
 [9]
 [10]
 [11]
 [12]
 [13]
 [14]
 [15]
 [16]
 [17]
 [18]
 [19]
 [20]
 [21]
 [22]
 [23]
 [24]
 [25]
 [26]
 [27]
 [28]
 [29]
 [30]
 [31]
 [32]
 [33]
 [34]
 [35]
 [36]
 [37]
 [38]
 [39]
 [40]
 [41]
 [42]
 [43]
 [44]
 [45]
 [46]
 [47]
 [48]
 [49]
 [50]
 [51]
 [52]
 [53]
 [54]
 [55]
 [56]
 [57]
 [58]
 [59]
 [60]
 [61]
 [62]
 [63]
 [64]
 [65]
 [66]
 [67]
 [68]
 [69]
 [70]
 [71]
 [72]
 [73]
 [74]
 [75]
 [76]
 [77]
 [78]
 [79]
 [80]
 [81]
 [82]
 [83]
 [84]
 [85]
 [86]
 [87]
 [88]
 [89]
 [90]
 [91]
 [92]
 [93]
 [94]
 [95]
 [96]
 [97]
 [98]
 [99]
 [100]

Query Match 88.4%; Score 1224; DB 2; Length 414;
 Best Local Similarity 87.3%; Pred. No. 1.8e-96;
 Matches 233; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

1 MEAHLATRLTNKVEPFPVLLISGHCALAVOGVSDPFLIGSKSLDIPAGMDLVKVAR 60
 148 MEAHLATRLTNKVEPFPVLLISGHCALAVOGVSDPFLIGSKSLDIPAGMDLVKVAR 207
 61 LSLIKHPECSTMSGKAI EHLAKOGNRFHDIKPELHAKNCDPFTGLOHTDKIMKK 120
 208 LSLIKHPECSTMSGKAI EHLAKOGNRFHDIKPELHAKNCDPFTGLOHTDKIMKK 267
 121 EKEGIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKORDLLPQNNAVLVASGV 180
 268 EKEGIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKORDLLPQNNAVLVASGV 327
 181 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 240
 328 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 387
 241 PKCPGLGVDSKEVGEASIKVPOLKMEI 267
 388 PKCPGLGVDSKEVGEASIKVPOLKMEI 414

RESULT 5
 09DON0 PRELIMINARY; PRT; 414 AA.

09DON0; PRELIMINARY; PRT; 414 AA.
 01-JUN-2001 (TREMBLrel. 17, Created)
 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610001M19 product:similar to PUTATIVE STALOGLYCOPROTEASE TYPE 2.

GN Name=Osegep11;
 OS Mus musculus (Mouse).
 OC Bkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P.; Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Mech. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20530913; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA The FANTOM Consortium,
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P.; Shibata Y.; Hayatsu N.; Sugahara Y.; Shibata K.; Itoh M.;
 RA Kono H.; Okazaki Y.; Muramatsu M.; Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper discovered cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K.; Itoh M.; Aizawa K.; Nagaoka S.; Sasaki N.; Carninci P.;
 RA Kono H.; Akiyama Y.; Nishi K.; Katsunai T.; Tashiro H.; Itoh M.;
 RA Suni N.; Ishii Y.; Nakamura S.; Hazama M.; Nishibe T.; Harada A.;
 RA Yamamoto R.; Matsunoto H.; Sakaguchi S.; Ikegami T.; Kashiwagi K.;
 RA Fujiwaka S.; Inoue K.; Togawa Y.; Izawa M.; Ohara E.; Watanabe M.;
 RA Yoneda Y.; Ishikawa T.; Ozawa K.; Tanaka T.; Matsura S.; Kawai J.;
 RA Okazaki Y.; Muramatsu M.; Inoue Y.; Kira A.; Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multichannel sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA Adachi J.; Aizawa K.; Akahira S.; Akimura T.; Arai A.; Aono H.;
 RA Arakawa T.; Bono H.; Carninci P.; Fukuda S.; Fukunishi Y.; Furuno M.;
 RA Hanagaki T.; Hara A.; Hayatsu N.; Hiramoto K.; Hiraoka T.; Hori F.;
 RA Imotani K.; Ishii Y.; Itoh M.; Izawa M.; Kasukawa T.; Kato H.;
 RA Kawai J.; Kojima Y.; Kono H.; Kouda M.; Koya S.; Kuribara C.;
 RA Matsuyama T.; Miyazaki A.; Nishi K.; Nomura K.; Numazaki R.; Ohno M.;
 RA Okazaki Y.; Okido T.; Owa C.; Saito H.; Saito R.; Sakai C.; Sakai K.;
 RA Sano H.; Sasaki D.; Shibata K.; Shibata Y.; Shinagawa A.; Shiraki T.;
 RA Sogabe Y.; Suzuki H.; Tagami M.; Tagawa A.; Takahashi F.; Tanaka T.;
 RA Tejima Y.; Toya T.; Yamamatsu T.; Yasunishi A.; Yoshida K.; Yoshino M.;
 RA Muramatsu M.; Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AK011265; BAB27506.1;
 DR MEROPS: M22.004;
 DR MGD; MG11919335; Osegep11.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.

DR InterPro; IPR009180; Pept M22_Osialg1.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialgic_ptide; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR TrEMBL; TrEMBL00329; gcp; 1.
 KW Peptidase.
 SQ SEQUENCE 414 AA; 44999 MW; 999BC689944DDB24 CRC64;
 Query Match 87.8%; Score 1216; DB 2; Length 414;
 Best Local Similarity 86.9%; Pred. No. 8.9e-96;
 Matches 232; Conservative 14; Mismatches 21; Indels 0; Gaps 0;
 QY 1 MEAHLATRLTKVPEPFVLLISGSHCLTAVGVSDPLLGKSLDIPAGMDPVAR 60
 DB 148 MEAHLATRLTKVPEPFVLLISGSHCLTAVGVSDPLLGKSLDIPAGMDPVAR 207
 QY 61 LSLIKHPECSWTSGGKALEHLAKQGNRPFDIKPPLHAKNCFSTGLQHYTDKIMRK 120
 DB 208 LSLIKHPECSWTSGGKALEHLAKQGNRPFDIKPPLHAKNCFSTGLQHYTDKIMRK 267
 QY 121 EKEBGEIKQIISADIAATVQHTMACHLVTRTHAILFCQRIIDLPONNAVYASGV 180
 DB 268 EKEBGEIKQIISADIAATVQHTMACHLVTRTHAILFCQRIIDLPONNAVYASGV 327
 QY 161 ASNFYIRALRTLTMACTCTCPPLCTDNGIMIANNGIERLRAGGLIDIGIRYE 240
 DB 328 ASNFYIRALRTLTMACTCTCPPLCTDNGIMIANNGIERLRAGGLIDIGIRYE 387
 QY 241 PKCPPLGVDSKEVGEASIKVPOLKMEI 267
 DB 368 PKCPPLGVDSKEVGEASIKVPOLKMEI 414
 RESULT 6
 O6AYN7 PRELIMINARY; PRT; 467 AA.
 AC O6AYN7;
 DT 25-OCT-2004 (TREMBLrel. 28, Created)
 DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCB1_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;
 RA Klausner R.D.; Collins F.S.; Wagner L.; Shenmen C.M.; Schuler G.D.;
 RA Altshul S.F.; Zeeberg B.; Buetow K.H.; Scheffer C.F.; Bhat N.K.;
 RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Heish F.;
 RA Diatchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;
 RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Caeavart T.L.; Scheetz T.E.;
 RA Brownstein M.J.; Uebli T.B.; Toshiyuki S.; Carninci P.; Prange C.;
 RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullaly S.J.;
 RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;
 RA Richards S.; Morley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W.;
 RA Villalón D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;
 RA Rahny J.; Helton E.; Kettelman M.; Madan A.; Rodriguez S.; Sanchez A.;
 RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;
 RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;
 RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield J.S.;
 RA Krzywinski M.I.; Skalska U.; Smalios D.E.; Scherch A.; Schein J.E.;
 RA Jones S.J.; Marz M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;

RA Director MGC Project;
 RL Submitted (Aug-2004) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC078974; AAT8974.1; -
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR009095; Peptidase M22.
 DR InterPro: IPR009180; Pept M22 O-sialyl.
 DR InterPro: IPR002016; Peroxidase.
 DR Pfam: PF00814; Peptidase M22; 1.
 DR PIRSF: PIRSF004537; O-sialyl_glyc_pptd; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR ProDom: PD002367; Peptidase M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR Hypothetical protein.
 KW NON_TER
 FT SEQUENCE 467 AA; 50799 MW; 474E18B1959B8AC0 CRC64;
 SQ
 Query Match 80.6%; Score 1116; DB 2; Length 467;
 Best Local Similarity 88.4%; Pred. No. 3.8e-87;
 Matches 214; Conservative 10; Mismatches 18; Indels 0; Gaps 0;
 Db 1 MEAHALTIRLTNKVPPPLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVKVAR 60
 148 MEAHALTIRLTNKVPPPLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVKVAR 207
 QY 61 LSLIKPECSMGSGKAI EHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDTKIIMKK 120
 Db 208 LSLIKPECSMGSGKAI EHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDTKIIMKK 267
 QY 121 EKEGIEKQILSSAADIAATVQHTMACLVKTRTRALIFCKORDLLPQNNAVLVASGV 180
 Db 268 EKEGIEKQILSSAADIAATVQHTMACLVKTRTRALIFCKORDLLPQNNAVLVASGV 327
 QY 181 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANNGIERLAGILHDIEGIRYE 240
 Db 328 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANNGIERLAGILHDIEGIRYE 387
 QY 241 PK 242
 Db 388 PK 389
 RESULT 7
 Q96NH5 PRELIMINARY; PRT; 364 AA.
 AC Q96NH5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ30879.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Magatsuna M., Shiratori A.,
 RA Muraaki H., Hoshino T., Kaku Y., Kodaïra H., Kondo H., Sugawara M.,
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki K.,
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Musashino K., Yunki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsumura H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Momiyama H., Saich N., Takami S., Terashima Y., Suuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togsashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Sato H., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi Y., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 DR EMBL: AK055441; BAB70923.1; -
 DR MEROPS: M22.004; -
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR009095; Peptidase M22.
 DR InterPro: IPR009180; Pept M22 O-sialyl.
 DR Pfam: PF00814; Peptidase M22; 1.
 DR PIRSF: PIRSF004537; O-sialyl_glyc_pptd; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR ProDom: PD002367; Peptidase M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 KW PROTEASE.
 SQ SEQUENCE 364 AA; 39528 MW; E0B605A07D0ECAD6 CRC64;
 Query Match 79.8%; Score 1105; DB 2; Length 364;
 Best Local Similarity 98.6%; Pred. No. 2.5e-86;
 Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Db 1 MEAHALTIRLTNKVPPPLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVKVAR 60
 148 MEAHALTIRLTNKVPPPLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVKVAR 207
 QY 61 LSLIKPECSMGSGKAI EHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDTKIIMKK 120
 Db 208 LSLIKPECSMGSGKAI EHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDTKIIMKK 267
 QY 121 EKEGIEKQILSSAADIAATVQHTMACLVKTRTRALIFCKORDLLPQNNAVLVASGV 180
 Db 268 EKEGIEKQILSSAADIAATVQHTMACLVKTRTRALIFCKORDLLPQNNAVLVASGV 327
 QY 181 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIA 217
 Db 328 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIA 364
 RESULT 8
 Q8JFW3 PRELIMINARY; PRT; 404 AA.
 AC Q8JFW3;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE SI:0221013.4 (Novel glycoprotease).
 GN Name=dz72814.6;
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NC NCB1_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Babage A.;
 RL Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.

DR EMBL; AL591593; CAD43471.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF04537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
KM Protease.
SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737E35 CRC64;
Query Match 62.0%; Score 859; DB 2; Length 404;
Best Local Similarity 62.3%; Pred. No. 3,4e-65;
Matches 165; Conservative 38; Mismatches 62; Indels 0; Gaps 0;
QY 1 MEAAHLLTRLTNKVPEPPLVLLISGGHCLALVQGVSDFLIGKSLDIAPGMDLVAVAR 60
DB MEAAHLLTRLMHPDPPLVLLVSGHSLALAKGIDFLLIGQLDEAAGDTLKIARR 195
QY 61 LSLIKHECSTMSGKAIHLAKOGNRFHPDIKPLHAKNCDFFSTGLQHTVDKIMKK 120
DB LSLNNHPECSTLSSGGAIERLAKSGDRLAFHIFPMGQNYDCNFSFAGLRTOITGAINK 255
QY 121 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCRKORDLPQNNAVLVASGV 180
DB EKEEGVEKGQFLSCVKDIAAASQHTVASHLAKRTRAILFCRSKGLLDEQNPFLIVSGGV 315
QY 181 ASNFYIRRALIELTNATQCTLLCPPELCTDNGIMIANNGIERLRAGLILHDIGIRYE 240
DB ASNFYIRQLIKITDAGTGLHLCPSPKCTDNGVMIANNGIERLKQGGKILSYSEVSVE 375
QY 241 PKCPILGVDISKVEGASIVPOLKM 265
DB PKAPPLGDLITSEVKEAIVKPKKL 400
RESULT 9
Q8JFR7 PRELIMINARY; PRT; 404 AA.
AC Q8JFR7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE SI:d72B14.6 (Novel glycoprotease).
GN Name=SI:d72B14.6;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark G.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL672217; CAD43443.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF04537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
KM Protease.

SQ SEQUENCE 404 AA; 44027 MW; 6FB98653A651860F CRC64;
Query Match 61.7%; Score 855; DB 2; Length 404;
Best Local Similarity 61.9%; Pred. No. 7.6e-65;
Matches 164; Conservative 39; Mismatches 62; Indels 0; Gaps 0;
QY 1 MEAAHLLTRLTNKVPEPPLVLLISGGHCLALVQGVSDFLIGKSLDIAPGMDLVAVAR 60
DB MEAAHLLTRLMHPDPPLVLLVSGHSLALAKGIDFLLIGQLDEAAGDTLKIARR 195
QY 61 LSLIKHECSTMSGKAIHLAKOGNRFHPDIKPLHAKNCDFFSTGLQHTVDKIMKK 120
DB LSLNNHPECSTLSSGGAIERLAKSGDRLAFHIFPMGQNYDCNFSFAGLRTOITGAINK 255
QY 121 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCRKORDLPQNNAVLVASGV 180
DB EKEEGVEKGQFLSCVKDIAAASQHTVASHLAKRTRAILFCRSKGLLDEQNPFLIVSGGV 315
QY 181 ASNFYIRRALIELTNATQCTLLCPPELCTDNGIMIANNGIERLRAGLILHDIGIRYE 240
DB ASNFYIRQLIKITDAGTGLHLCPSPKCTDNGVMIANNGIERLKQGGKILSYSEVSVE 375
QY 241 PKCPILGVDISKVEGASIVPOLKM 265
DB PKAPPLGDLITSEVKEAIVKPKKL 400
RESULT 10
Q7Q918 PRELIMINARY; PRT; 401 AA.
AC Q7Q918;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE AGCP14990 (Fragment).
GN Name=agCG46164; ORFNames=ENSGANG0000007922;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_Taxid=180454;
RN [1]
RP SEQUENCE FROM N.A.
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008900; EAA09387.1; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF04537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
FT NON TER 1
SQ SEQUENCE 401 AA; 43849 MW; CC9426723D3FD4F1 CRC64;
Query Match 32.7%; Score 453.5; DB 2; Length 401;
Best Local Similarity 38.4%; Pred. No. 1.9e-30;
Matches 106; Conservative 50; Mismatches 109; Indels 11; Gaps 4;
QY 1 MEAAHLLTRLTNKVPEPPLVLLISGGHCLALVQGVSDFLIGKSLDIAPGMDLVAVAR 60
DB MOAHLALMAWMTSTIPPLCLVSGHSLLVVESTARLLAGETLDDAPGALDKIARR 178
QY 61 LSLIKHECSTMSGKAIHLAKOG----NRFHPDIKPLHAKNCDFFSTGLQHTVDK 115
DB LKLRVAVAKYAKQMSGGAIERLAKGAKQTSAYVEPFL--PLSKYRRCQGSFAGLRKVTARR 236

RESULT	12
ID	Q9VWMD6
AC	Q9VWMD6
DT	01-MAY-2000 (TREMBLrel. 13, Created)
D7	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
D7	01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE	CGI42311.PA.
GN	ORFNames=CGI42311;
OS	Drosophila melanogaster (fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Nemoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxId=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA	George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.H., Blair R.G., Champe M., Pfeiffer B.D.,
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
RA	Abril J.F., Agagyan A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA	Balcer R.M., Basu A., Baxendale A.J., Bayraktaroglu I., Beasley E.M.,
RA	Bessert K.Y., Benos P.V., Bertana B.P., Bhargava D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brocktein P., Brotlier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., May A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doull L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Foster C., Gabrielian A.E., Gang N.Z., Gelbart W.M., Glaeser K.,
RA	Glocke A., Gong P., Gorelli J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.U., Hernandez J.R., Houck J.,
RA	Hoscht D., Houston K.A., Howell T.J., Wei M.H., Ibegwam C.,
RA	Jalali B., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., liang Y., Lin X.,
RA	Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Modarray C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Mutzy D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA	Reinert K., Remington K., Sanders R.D., Scheefler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA	Williams S.M., Woodgett M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA	Zheng R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA	Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of Drosophila melanogaster.";
RL	Science 287:2185-2195(2000).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22426055; PubMed=12537568;
RA	Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA	Ratel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA	George R.A., Hoskins R.A., Laverty T., Mutzy D.M., Nelson C.R.,
RA	Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA	Svirskas R., Tabot P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA	Weinstock G., Scherer S.B., Myers B.W., Gibbs R.A., Rubin G.M.,
RT	"Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT	melanogaster euchromatic genome sequence.";
RL	Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22426070; PubMed=12537573;
RA	Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA	Ratel S., Frise E., Wheeler D.A., Lewis S.B., Rubin G.M.,

RA Ashburner M., Celisner S.B.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 a genomic perspective."
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu U., Berman B.P.,
 RA Beltercourt B.R., Celisner S.B., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.B.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 systematic review."
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AEO03513; AAF49008.1; -
 DR IntAct; Q9VWD6; -
 DR FlyBase; FBgn0031060; CG14231.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF04537; Osialgl_ptcds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR SEQUENCE 409 AA; 45328 MW; 9797F667D155538 CRC64;
 SQ
 Query Match 30.0%; Score 415.5; DB 2; Length 409;
 Best Local Similarity 38.2%; Pred. No. 3.6e-27;
 Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;
 QY 1 MEAAALTRRLT--KVEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 58
 DB MEAAALTRRLT--KVEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 58
 QY 136 MEAAALTRRLT--KVEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 195
 DB MEAAALTRRLT--KVEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 195
 QY 59 RRLSLIHPECSTWMSGKAIHLAK--QGNRPHFDIKRPLHAKNCDSPFTGLQHYTDKI 117
 DB RRLSLIHPECSTWMSGKAIHLAK--QGNRPHFDIKRPLHAKNCDSPFTGLQHYTDKI 117
 QY 118 MKKEKEGIEKQGLLSNADIAATVQHTMAGLVKTRHAILFC--KORDLLPQNNAVLV 175
 DB MKKEKEGIEKQGLLSNADIAATVQHTMAGLVKTRHAILFC--KORDLLPQNNAVLV 175
 QY 256 RARRRARTRPDGVISNVCDFCAGLRVSNSHLMRTQRAIEYICLLPRHQLFGDIPPLIV 315
 DB RARRRARTRPDGVISNVCDFCAGLRVSNSHLMRTQRAIEYICLLPRHQLFGDIPPLIV 315
 QY 176 ASGCVASNFYIRALLETITNAQTCTLLCPRLCTDNGIMTAWNGIERLRAGLIADIE 235
 DB ASGCVASNFYIRALLETITNAQTCTLLCPRLCTDNGIMTAWNGIERLRAGLIADIE 235
 QY 316 MSGGVANNDAIYANIEHLAAGYGCSPFRPSKRYCSDNGMIMHGVDEL-----LOKKE 369
 DB MSGGVANNDAIYANIEHLAAGYGCSPFRPSKRYCSDNGMIMHGVDEL-----LOKKE 369
 QY 236 -GIRYEPKCPGLGVNISKEVGEA 256
 DB -GIRYEPKCPGLGVNISKEVGEA 256
 QY 370 ASTRYDYD--SIDIQSAGFA 388
 DB ASTRYDYD--SIDIQSAGFA 388
 RESULT 13
 022145 PRELIMINARY; PRT; 480 AA.
 AC 022145; Q8VWL2;
 DT 01-JUN-1998 (TREMBlrel. 05, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease

DE GCP1).
 GN Name=At2g45270; Synonyms=GCP1;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Mason T.M., Kertavage A.R., Adams M.D., Somerville C.R.,
 RA Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hauswirth K., Adamska I.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carinci P., Chen H., Cheuk R., Hayaishiaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narueeka M., Nguyen M., Palm C.D., Sakurai T., Satou M.,
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.W., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carinci P., Chen H., Cheuk R., Hayaishiaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narueeka M., Nguyen M., Palm C.D., Sakurai T., Satou M.,
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002387; AAB82636.2; -
 DR EMBL: AY024338; AAK00530.1; -
 DR EMBL: AY063864; AAL36220.1; -
 DR EMBL: AY117283; AAM51358.1; -
 DR PIR; B84888; B84888.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF04537; Osialgl_ptcds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR SEQUENCE 480 AA; 52995 MW; 20DD6A86ACC1FFAD CRC64;
 SQ
 Query Match 29.4%; Score 407.5; DB 2; Length 480;
 Best Local Similarity 37.0%; Pred. No. 2.1e-26;
 Matches 104; Conservative 37; Mismatches 93; Indels 47; Gaps 7;
 QY 1 MEAAALTRRLT--NKEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 59
 DB MEAAALTRRLT--NKEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 59
 QY 135 MEAAALTRRLT--NKEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 254
 DB MEAAALTRRLT--NKEFPVLVLSGGHCLALVQGVSDFLKSKSLDIAPGMDLKVX 254
 QY 60 RSLIKHPECSTWMSGKAIHLAKQGNRPHFDIKRPLHAKNCDSPFTGLQHYTDKI 119
 DB RSLIKHPECSTWMSGKAIHLAKQGNRPHFDIKRPLHAKNCDSPFTGLQHYTDKI 119
 QY 255 WIGLDMH-----RSSGPAVEELALSGDAKSVFNVPMKVKHCKCNFSGYAGLKTQVRLAIEA 309
 DB WIGLDMH-----RSSGPAVEELALSGDAKSVFNVPMKVKHCKCNFSGYAGLKTQVRLAIEA 309

```

QY 120 KEKEGIEKQIILSSA-----ADIAATVOHTMACHLVKRTTRAILFCRKORLLPQNN 171
DB 310 KE-----IDAKCPVSSATNEDBRNRADIAASFOVAVLHLEEKCEBIAIDALE---LEPSI 362
QY 172 AVLVASGVANFYTRIRALSLITNAQTCTLLCPPLRLCTDNGIMIANGLIERLAGLGL 231
DB 363 KHWIISGVANSKYRLRLNINIIVENKMLKLVCPPLSLCTDNGVWVAMTGLIEHFRVG--- 418
QY 232 HDIGIRYE-----PKCPGLVDISKVEGVA 256
DB 419 -----RYDPPPPATPEPDYDYDLRPRMPLGBEVAKGSEA 453

RESULT 14
ID 073H71 PRELIMINARY; PRT; 335 AA.
AC 073H71;
DT 05-JUL-2004 (TREMblrel. 27, Created)
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)
DE Peptidase, M22 family protein.
GN OrderedCusNames=WD0659;
OS Wolbachia pipiensis wMel.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Wolbachiae; Wolbachia.
OX NCBI_TaxId=66077;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15024419;
RA Wu M., Sun L.V., Vamathevan J.J., Riegler M., DeBoy R.T.,
RA Brownie J.C., McGraw E.A., Martin W., Esser C., Abmadinjad N.,
RA Wiegand C., Madupur R., Beaman M.J., Brinkac L.M., Daugherty S.C.,
RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,
RA Berry K.J., Young M.B., Uterback T.R., Weidman J.F., Nierman W.C.,
RA Paulsen I.T., Nelson K.E., Tettelin H., O'Neill S.L., Eisen J.A.,
RA "Phylogenomics of the reproductive parasite Wolbachia pipiensis wMel:
RT a streamlined genome overrurn by mobile genetic elements."
RL Plos Biol. 2:327-341(2004).
DR EMBL; AE017258; AA014395.1; -.
DR TIGR; WD0659; -.
DR GO; GO:0008450; F:0-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00905; Peptidase_M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglcptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
DR KW Complete proteome.
SQ SEQUENCE 335 AA; 36634 MW; F9194CDA58BC9390 CRC64;

Query Match 26.6%; Score 369; DB 2; Length 335;
Best Local Similarity 34.7%; Pred. No.2.8e-23;
Matches 84; Conservative 46; Mismatches 86; Indels 26; Gaps 5;

QY 1 MEAAATLRLTNKVFPPVLVLLISGHCILALVQGVSPFLIGSLDIAPGMDLVKAR 60
DB 112 LEAAVATLRLHEKVPFLVLLISGHCQFLIADQVCKITLGETLDSLGAPFKVAKM 171
QY 61 LSLIHPECSMTSGGKATLHAKOGNRFHDIKPLHLHAKNCDPSFTGLQHTVTKIMKK 120
DB 172 LGL-----SYPCGPLIEKLAKKNGTRFPLPRMIRSGCNFSFGIKTVAKMLVQBL 224
QY 121 EKERGIEKQIILSSAADIAATVOHTMACHLVKRTTRAILFCRKORLLPQNNAVVASGV 180
DB 225 KWSB-----QDVCVGASFOECISDILLDRVSNALIMASLNIKIND--PVTGCV 273
QY 161 ASNFYIRALSLITNAQTCTLLCPPLRLCTDNGIMIANGLIERLAGLGLHDIGIRYE 240
DB 274 NANNPLAEKIKOHIN--LNIFFPPNDLCTDNALMVGMTGIERLOKNT-----IDPLAFA 325

```

```

QY 241 PK 242
DB 326 PR 327

RESULT 15
ID 092LH8 PRELIMINARY; PRT; 360 AA.
AC 092LH8;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).
GN ORFNames=SMC03230;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxId=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kies E., Lelaure V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.,
RA "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591792; CAC47657.1; -.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0008450; F:0-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00905; Peptidase_M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglcptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
DR KW Complete proteome; Hydrolase.
SQ SEQUENCE 360 AA; 37906 MW; A07F946B562EA86 CRC64;

Query Match 25.7%; Score 356.5; DB 2; Length 360;
Best Local Similarity 35.6%; Pred. No.3.5e-22;
Matches 95; Conservative 38; Mismatches 91; Indels 43; Gaps 7;

QY 1 MEAAATLRLTNKVFPPVLVLLISGHCILALVQGVSDPLIGSLDIAPGMDLVKAR 60
DB 116 LEHALTLRLTGLTFPPVLMVLVSGHTQLIIVKGGEYERWGTITDDALGEAFDTAKL 175
QY 61 LSLIHPECSMTSGGKATLHAKOGNRFHDIKPLHLHAKNCDPSFTGLQ----- 110
DB 176 LGL-PP-----GGPAVERAAQGAERFDPPLVGGARLDPSGLTAVRQAQSL 228
QY 111 -HTVDKIMKKEKEGIEKQIILSSAADIAATVOHTMACHLVKRTTRAILFCRKORLLPQ 169
DB 229 GAVTQDI-----ADVCAFOAIRISITLDRVGRGLKRRP-ADFAV 269
QY 170 NNAVIVASGVANFYTRIRALSLITNAQTCTLLCPPLRLCTDNGIMIANGLIERLAGL 229
DB 270 DQPALVIVAGVANAQTLRRTTQSLDEHGFRTAPPLAQCTDNAAAMIWAGAERLAGL 328
QY 230 ILHDIEGIRYEPRK--PLGVDISKVEG 254
DB 329 ---PADGLDAAPRSRWPIDSRKALIG 352

Search completed: February 16, 2005, 13:08:40
UOD time : 47.2739 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 08:37:51 ; Search time 48.0598 Seconds
(without alignment)
1923.349 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVQGVSDPFLILGKSLDI.....DISKEVGRASIKVPQLKMEI 239

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
1: A_Geneseq_16Dec04:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003as:*
8: geneseqp2003bs:*
9: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	267	5	ABG96487 Novel hum
2	1240	100.0	414	5	ABG96478 Novel hum
3	1240	100.0	414	5	ABB05481 Human O-8
4	1240	100.0	414	6	ABJ26654 Human pro
5	1203	97.0	414	5	AAE29234 Human gly
6	1203	97.0	414	6	ABG71161 Human gly
7	1203	97.0	414	6	ABU09569 Human gly
8	960	77.4	364	6	ADA54471 Human gly
9	332.5	26.8	409	4	ABB69133 Drosophila
10	326	26.3	463	3	AAV52216 Arabidops
11	319	25.7	245	3	AAV52216 Arabidops
12	319	25.7	439	3	AAV52216 Arabidops
13	319	25.7	444	3	AAV52216 Arabidops
14	257	20.7	382	6	ABU22934 Protein e
15	248.5	20.0	348	6	ABU16575 Protein e
16	247	19.9	312	6	ABU16575 Protein e
17	247	19.9	350	8	ADL05040 M. catara
18	241	19.4	343	6	ABU39303 Protein e
19	239	19.3	251	5	AAE31054 Ehrlichia
20	231	18.6	341	4	AAU36205 Pseudomon
21	231	18.6	341	6	ABU38276 Protein e
22	231	18.6	341	7	ABG73342 P aerugin
23	231	18.6	401	7	ABO68626 Pseudomon
24	229	18.5	341	7	ADG73344 P aerugin
25	224	18.1	342	3	AAV52202 Haemophil

26	224	18.1	342	4	AAU35450 Haemophil
27	224	18.1	342	6	ABU30280 Protein e
28	213.5	17.2	347	4	ABE61112 Drosophila
29	213	17.2	337	4	AAU38187 Salmonella
30	213	17.2	337	6	ABU47561 Protein e
31	212	17.1	340	6	ABU40514 Protein e
32	212	17.1	357	7	ADFO6228 Bacterial
33	210	16.9	337	6	ABU50237 Protein e
34	209	16.9	326	4	AAE96423 Putative
35	208	16.8	335	6	ABU27480 Protein e
36	207.5	16.7	341	6	ABU40069 Protein e
37	206	16.6	337	3	AAV52204 Escherich
38	206	16.6	337	4	AAU34711 E. coli c
39	206	16.6	337	7	ABU28771 Protein e
40	205	16.5	343	7	ABO62704 Klebsiella
41	201.5	16.2	421	5	ABG96491 Novel hum
42	200	16.1	325	2	AAE26325 Glycoprot
43	200	16.1	325	3	AAV52203 Pasteurel
44	199.5	16.1	350	6	ABU25741 Protein e
45	199	16.0	354	6	ABU37844 Protein e

ALIGNMENTS

RESULT 1
ID ABG96487 standard; protein, 267 AA.
AC ABG96487;
XX
DT 11-DEC-2002 (first entry)
XX
DE Novel human metalloprotease MPI fragment #1.
XX
KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder.
XX
OS Homo sapiens.
XX
PN WO200272751-A2.
XX
PD 19-SEP-2002.
XX
PP 05-FEB-2002; 2002WO-US003353.
XX
PR 05-FEB-2001; 2001US-0266518P.
XX
PR 10-APR-2001; 2001US-0282814P.
XX
(BRIM) BRISTOL-MYERS SQUIBB CO.
XX
Chen J, Feder J, Nelson TC, Duclos F, Krystek S,
XX
WPI, 2002-72329/78.
XX
DR N-PSDB; ABS76639.
XX
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.
XX
PS Claim 5; Page 29; 473p; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or

CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloprotease MP1 protein

XX Sequence 267 AA;

Query Match 100.0%; Score 1240; DB 5; Length 267;

Best Local Similarity 100.0%; Pred. No. 7,9e-131; Mismatches 0; Gaps 0;

Matches 239; Conservative 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDFLILGKSLDIAPGMDLVKVARLSLIKHECSTMSGKAIIEHLAKQGNRF 60
 DB 29 LIALVQGVSDFLILGKSLDIAPGMDLVKVARLSLIKHECSTMSGKAIIEHLAKQGNRF 88
 QY 61 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMAC 120
 DB 89 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMAC 148
 QY 121 HLVKRTHRAILFCCKORDILPQNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPRL 180
 DB 149 HLVKRTHRAILFCCKORDILPQNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPRL 208
 QY 181 CTDNGIMIAMNGIERLRAGILGILHDIEGIRYEPKPCPLGVDISKVEGASIKVPOLKMEI 239
 DB 209 CTDNGIMIAMNGIERLRAGILGILHDIEGIRYEPKPCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 2

ABG96478 ID ABG96478 standard; protein; 414 AA.

XX ABG96478;

DT 11-DEC-2002 (first entry)

XX Novel human metalloprotease MP1.

XX Metalloprotease: MP-1; immune disorder; glutamate transport; cancer;
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 KW neurological disorder.

OS Homo sapiens.

XX WO200272751-A2.

XX 19-SEP-2002.

XX 05-FEB-2002; 2002WO-US003353.

XX 05-FEB-2001; 2001US-0266518P.
 PR 10-APR-2001; 2001US-0282814P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
 XX MPI: 2002-723329/78.
 DR N-PSDB; ABS76635.

XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,
 PT treating, or ameliorating diseases associated with aberrant
 PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
 PT neurological disorders.

XX Claim 5; Fig 1A-C; 473pp; English.

PS The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloprotease MP1 protein

XX Sequence 414 AA;

Query Match 100.0%; Score 1240; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.5e-130; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDFLILGKSLDIAPGMDLVKVARLSLIKHECSTMSGKAIIEHLAKQGNRF 60
 DB 176 LIALVQGVSDFLILGKSLDIAPGMDLVKVARLSLIKHECSTMSGKAIIEHLAKQGNRF 225
 QY 61 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMAC 120
 DB 236 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMAC 295
 QY 121 HLVKRTHRAILFCCKORDILPQNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPRL 180
 DB 236 HLVKRTHRAILFCCKORDILPQNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPRL 355
 QY 181 CTDNGIMIAMNGIERLRAGILGILHDIEGIRYEPKPCPLGVDISKVEGASIKVPOLKMEI 239
 DB 356 CTDNGIMIAMNGIERLRAGILGILHDIEGIRYEPKPCPLGVDISKVEGASIKVPOLKMEI 414

RESULT 3

ABB05481 ID ABB05481 standard; protein; 414 AA.

XX ABB05481;

DT 19-APR-2002 (first entry).

XX Human O-6-allylglycoproteinase-like protein SEQ ID NO:2.

XX Human; O-6-allylglycoproteinase-like protein; OSGPLP; enzyme.

XX Homo sapiens.

XX CN1318550-A.

XX 24-OCT-2001.

XX 19-APR-2000; 2000CN-00106834.

PR 19-APR-2000; 2000CN-00106834.
XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
XX Mao Y, Xie Y;
XX WPI; 2002-115090/16.
DR N-PSDB; ABA93268.
XX
PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
PT for diagnosing, preventing and treating related diseases.
XX
PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.
XX
CC The present sequence represents human O-sialoglycoproteinase-like protein
CC (OSGPRP). The present invention also describes: (1) the preparation of
CC the OSGPRP protein; (2) applying the OSGPRP protein in diagnosis; (3) the
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPRP
CC protein in screening its agonist, excitomotor and inhibitor and preparing
CC an antibody against the OSGPRP protein; and (5) the use of the OSGPRP
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors
CC and antibodies in treating diseases related to the abnormal OSGPRP gene
CC and in preparing the medicine composite for the treatment
XX
XX Sequence 414 AA;
XX
Query Match 100.0%; Score 1240; DB 5; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLAIVQGVSDPFLILGKSLDIAPGMDLDRVARRSLIKHPBCSTWSGKAIIEHLAKGNRF 60
DB 176 LLAIVQGVSDPFLILGKSLDIAPGMDLDRVARRSLIKHPBCSTWSGKAIIEHLAKGNRF 235
QY 61 HPDIKPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 236 HPDIKPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKGQILSSAADIAATVQHTMAC 295
QY 121 HLKRTTRAILFCQKRDILLPQNNAVLVASGVASNFYIRRALEITNATQCTLLCPPPRL 180
DB 296 HLKRTTRAILFCQKRDILLPQNNAVLVASGVASNFYIRRALEITNATQCTLLCPPPRL 355
QY 181 CTDNGIMTAMNGIERLRAGILGILHDIIGIRYRKPCLGVDISKVEGASIKVPQLKMEI 239
DB 356 CTDNGIMTAMNGIERLRAGILGILHDIIGIRYRKPCLGVDISKVEGASIKVPQLKMEI 414
RESULT 4
ABJ26654
ID ABJ26654 standard; protein; 414 AA.
XX
AC ABJ26654;
XX
DT 01-MAY-2003 (first entry)
XX
DE Human protein modification + maintenance molecule protein SEQ ID No 8.
XX
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
XX antitumor; hepatotropic; gynecological; antibacterial; vinicide;
XX protozoacide; antiparasitic; cell proliferative disease; PMOD;
XX protein modification and maintenance molecule; immunogenic fragment;
XX cancer; autoimmune; inflammatory disease; neurological disorder;
XX gastrointestinal; developmental; vesicle trafficking disorder; infection;
XX protein-protein interaction; drug-target interaction;
XX gene expression profile; human.
OS Homo sapiens.
XX
XX NC030306644-42.
XX
XX 03-JAN-2003.
XX

PF 18-JUN-2002; 2002MO-US019360.
XX
XX 22-JUN-2001; 2001US-0300508P.
PR 06-JUL-2001; 2001US-0303445P.
PR 13-JUL-2001; 2001US-0305405P.
PR 09-AUG-2001; 2001US-0311442P.
PR 24-AUG-2001; 2001US-0314821P.
PR 29-AUG-2001; 2001US-0315992P.
XX 03-MAY-2002; 2002US-0378205P.
XX
XX (INCYTE GENOMICS INC.
XX
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BW,
XX Warren BA, Ison CH, Honchall CD, Nguyen DB, Lu DAM, Lee EA, Yue H;
XX Foreythe JF, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang Y;
XX Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Boroweky ML, Yao MG;
XX Walla NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK;
XX Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebardjian Y;
XX WPI; 2003-184039/18.
DR N-PSDB; ABR23207.
XX
XX New isolated human PMOD polypeptide and polynucleotide, useful for
XX diagnosing, treating and preventing diseases or conditions associated
XX with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
XX infections.
XX
XX Claim 63; Page 182-183; 225pp; English.
XX
XX The invention relates to an isolated polypeptide comprising: any of 28
XX sequences of 48-1256 amino acids; a natural amino acid sequence at least
XX 90% identical to the 28 amino acid sequences, 94% identical to a sequence
XX of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
XX acids, or 97% identical to a sequence of 242 amino acids, all given in
XX the specification; or a biologically active or immunogenic fragment of
XX the isolated polypeptide. The polypeptides and polynucleotides are useful
XX in diagnosing, treating and preventing diseases or conditions associated
XX with the decreased expression of protein modification and maintenance
XX molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
XX atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
XX allergies), neurological disorders (e.g. stroke, Parkinson's disease,
XX epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
XX endometriosis), developmental, vesicle trafficking disorders, and
XX infections (e.g. bacterial, viral, parasitic, protozoal). These are also
XX useful in assessing the effects of exogenous compounds on the expression
XX of nucleic acid and amino acid sequences of PMOD. The PMOD or its
XX fragments are useful in screening compounds for effectiveness as agonist
XX or antagonist of the polypeptides, or in altering the expression of the
XX target polynucleotide and compounds that specifically bind to or modulate
XX the activity of the polypeptide. The microarray is useful in monitoring
XX or measuring protein-protein interactions, drug-target interactions, and
XX gene expression profiles. This sequence represents a human PMOD protein
XX of the invention
XX
XX Sequence 414 AA;
XX
Query Match 100.0%; Score 1240; DB 6; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLAIVQGVSDPFLILGKSLDIAPGMDLDRVARRSLIKHPBCSTWSGKAIIEHLAKGNRF 60
DB 176 LLAIVQGVSDPFLILGKSLDIAPGMDLDRVARRSLIKHPBCSTWSGKAIIEHLAKGNRF 235
QY 61 HPDIKPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 236 HPDIKPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKGQILSSAADIAATVQHTMAC 295
QY 121 HLKRTTRAILFCQKRDILLPQNNAVLVASGVASNFYIRRALEITNATQCTLLCPPPRL 180
DB 296 HLKRTTRAILFCQKRDILLPQNNAVLVASGVASNFYIRRALEITNATQCTLLCPPPRL 355
QY 181 CTDNGIMTAMNGIERLRAGILGILHDIIGIRYRKPCLGVDISKVEGASIKVPQLKMEI 239
DB 356 CTDNGIMTAMNGIERLRAGILGILHDIIGIRYRKPCLGVDISKVEGASIKVPQLKMEI 239


```
Db      356 CTDNGIMIAMNGIERLRAGILHDIEGIRYBPKCPGLVDISKVEASIKVQLKMEI 414
|||||
RESULT 5
AAE29234
ID      AAE29234 standard; protein; 414 AA.
XX
AC      AAE29234;
XX
DT      27-JAN-2003 (first entry)
XX
DE      Human glycoprotease 28472 protein.
XX
KW      Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
KW      7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;
KW      rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;
KW      hyperextension; ischaemic heart disease; obesity; myocardial infarction;
KW      endothelial cell disorder; Grave's disease; psoriasis; brain disorders;
KW      Parkinson's disease; Alzheimer's disease; hematopoietic disorder;
KW      cerebral oedema; metabolic disorder; liver disorder; platelet disorder;
KW      chromosome mapping; tissue typing; gene therapy; neuroprotective;
KW      cytosolic; anorectic; cardiatic; haemostatic.
XX
OS      Homo sapiens.
XX
FH      Key
FT      Domain
FT      Location/Qualifiers
FT      1..108
FT      /note= "Non-transmembrane domain; N-terminal cytoplasmic
FT      domain"
FT      38..369
FT      /note= "Endopeptidase O-sialoglycoprotein hydrolase
FT      metalloprotease zinc glycoprotease sialoglycoprotease
FT      domain"
FT      109..132
FT      /note= "Transmembrane domain"
FT      133..164
FT      /note= "Non-transmembrane domain; non-cytoplasmic loop"
FT      138..152
FT      /note= "Glycoprotease domain"
FT      165..189
FT      /note= "Transmembrane domain"
FT      190..316
FT      /note= "Non-transmembrane domain; cytoplasmic domain"
FT      317..333
FT      /note= "Transmembrane domain"
FT      334..414
FT      /note= "Non-transmembrane domain"
FT      374..414
FT      /note= "Sialoglycoprotease type domain"
XX
XX      WO200274960-A2.
XX
XX      26-SEP-2002.
XX
XX      08-NOV-2001; 2001WO-US051427.
XX
XX      08-NOV-2000; 2000US-0246768P.
XX      08-NOV-2000; 2000US-0246772P.
XX      15-NOV-2000; 2000US-0249185P.
XX
XX      (MIL-) MILLENNIUM PHARM INC.
XX
XX      LeiJy KR, Kapeller-Libermann R, Glucksmann M;
XX      WPI: 2002-759898/82.
XX      N-PSDB; AAD46856.
XX
XX      New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
XX      useful for diagnosing and treating cancer, immune, cardiovascular,
XX      hematopoietic, brain, pain, metabolic, liver or platelet disorders, and
XX      in pharmacogenomics.
```

```
PS      Claim 1; Fig 8; 178pp; English.
XX
XX      The present invention relates to novel 38650, 28472, 5495, 65507, 81588
XX      or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
XX      protease or seven transmembrane domain (7TM) receptor family members.
XX      Sequences of the invention are useful in diagnosing and treating cancer
XX      or aberrant cellular proliferation and/or differentiation (e.g. colon or
XX      lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
XX      arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
XX      hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
XX      myocardial infarction, thrombus) including endothelial cell disorders
XX      (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
XX      disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
XX      pain and metabolic disorders (e.g. obesity), liver disorders or platelet
XX      disorders. They are also useful in screening assays, predictive medicine
XX      (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
XX      and pharmacogenetics) and prophylactic and therapeutic methods. The
XX      nucleic acids may also be used in chromosome mapping, tissue typing and
XX      forensic biology and as surrogate markers. Sequences of the invention are
XX      also used in gene therapy. The present sequence is human glycoprotease
XX      28472 protein
XX
SQ      Sequence 414 AA;
XX
Query Match      97.0%; Score 1203; DB 5; Length 414;
Best Local Similarity 97.1%; Pred. No. 2,3e-126;
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
XX
QY      1 LIALVGVSDPFLILGKSLDIAPGDMIDKVARRLSLIKHEPCTSMGCKAI EHLAKGNRF 60
DB      176 LIALVGVSDPFLILGKSLDIAPGDMIDKVARRLSLIKHEPCTSMGCKAI EHLAKGNRF 235
QY      61 HPDIKPRPLHANPCPSFPGLOHYTDKIMKEKEGKIGKQILSSAADIAATVQHTMAC 120
DB      236 HPDIKPRPLHANPCPSFPGLOHYTDKIMKEKEGKIGKQILSSAADIAATVQHTMAC 295
QY      121 HLVKRTHRALIFCKORDLIPONNAVIVASGVASNFYIRRALIILNATQCTLLCPPEPL 180
DB      296 HLVKRTHRALIFCKORDLIPONNAVIVASGVASNFYIRRALIILNATQCTLLCPPEPL 355
QY      181 CTDNGIMIAMNGIERLRAGILHDIEGIRYBPKCPGLVDISKVEASIKVQLKMEI 239
DB      356 CTDNGIMIAMNGIERLRAGILHDIEGIRYBPKCPGLVDISKVEASIKVQLKMEI 414
XX
RESULT 6
ABG71161
ID      ABG71161 standard; protein; 414 AA.
XX
XX      ABG71161;
XX
XX      30-JAN-2003 (first entry)
XX
XX      Novel human glycoprotease 28472.
XX
XX      Cancer; aberrant cell proliferation; aberrant cell differentiation;
XX      breast cancer; ovarian cancer; prostate cancer; colon cancer;
XX      lung cancer; immune disorder; heart disorder; cardiovascular disorder;
XX      endocheial disorder; hematopoietic disorder; blood vessel disorder;
XX      brain disorder; pain; metabolic disorder; liver disorder; diabetes;
XX      platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;
XX      autoimmune disorder; hypertension; atherosclerosis; heart failure;
XX      myocardial infarction; ischaemic heart disease; Crohn's disease;
XX      Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
XX      cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;
XX      Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.
XX
OS      Homo sapiens.
XX
FH      Key
FT      Domain
FT      Location/Qualifiers
FT      138..152
FT      /label= Glycoprotease_domain
```


XX MO200277233-A2.
 XX 03-OCT-2002.
 XX
 PF 08-NOV-2001; 2001MO-US046724.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (MILL-) MILENNIUM PHARM INC.
 XX
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
 PI WPI; 2003-029318/02.
 DR N-PSDB; ABS57020.
 XX
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
 PT hypertension.
 XX
 PS Claim 4; Fig 8A-B; 178bp; English.
 XX
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
 CC sequences that encode a human seven transmembrane domain (7TM). The
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
 CC sequences are useful for diagnosing, preventing or treating a subject
 CC with or at risk of developing a disorder, e.g. cancer or aberrant
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
 CC prostate, colon or lung cancer), immune disorders, heart disorders,
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,
 CC liver disorders or platelet disorders. These disorders include carcinoma,
 CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,
 CC ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
 CC cachexia or diabetes. This is the amino acid sequence of the novel human
 CC glycoprotease 28472
 XX
 SQ Sequence 414 AA:
 Query Match 97.0%; Score 1203; DB 6; Length 414;
 Best Local Similarity 97.1%; Pred. No. 2.3e-126;
 Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 LLAIVQGVSDPFLLLKSLIDIAFGMDLVKVARLSLIKPECSWTSGGKAIETHAKQGRF 60
 DB 176 LLAIVQGVSDPFLLLKSLIDIAFGMDLVKVARLSLIKPECSWTSGGKAIETHAKQGRF 235
 QY 61 HPDIKPLHAKNCDPSFTGLQHTVDKTIIMKKKEGIEKKQIILSSADIAATVQHTMAC 120
 DB 236 HPDIKPLHAKNCDPSFTGLQHTVDKNNENKQEBEGIEKKQIILSSADIAATVQHTMAC 295
 QY 121 HLYKRTTRALLFCQKORLLPONNAVVAASGVASNPYRRLALLETITNATTOCTLLCPPRRL 180
 DB 296 HLYKRTTRALLFCQKORLLPONNAVVAASGVASNPYRRLALLETITNATTOCTLLCPPRRL 355
 QY 181 CTONGIMIANNGIERLRAGLIIHIDIGIRYEPKPCIGVDISKVGBASIVPOLKKEI 239
 DB 356 CTONGIMIANNGIERLRAGLIIHIDIGIRYEPKPCIGVDISKVGBASIVPOLKKEI 414

AC ABU09569;
 XX
 DT 08-JUL-2003 (first entry)
 XX
 DE Human glycoprotease encoded by cDNA 28472.
 XX
 KW Human; enzyme; cancer; aberrant cellular proliferation; differentiation;
 KW immune disorders; heart disorder; brain disorder;
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;
 KW haematopoietic disorder; blood vessel disorder; metabolic disorder;
 KW liver disorder; platelet disorder; glycoprotease.
 XX
 OS Homo sapiens.
 XX
 PN US2003009017-A1.
 XX
 PD 09-JAN-2003.
 XX
 PF 08-NOV-2001; 2001US-00012140.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (LEIBY) LEIBY K R.
 PA (KAPPE) KAPPELLER-LIBERMANN R.
 PA (GLUC) GLUCKSMANN M A.
 XX
 PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;
 PI WPI; 2003-428888/40.
 DR N-PSDB; ACA60887.
 XX
 CC The invention relates to an isolated 38650 (encoding adenosine
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
 CC a sequence which is at least 60% identical to the six nucleic acids or
 CC their open reading frames, fragments of at least 15 nucleotides,
 CC naturally occurring variants, or a DNA insert of the plasmid deposited
 CC with the American Type Culture Collection as Accession No. not defined in
 CC the specification, which encodes the amino acid sequence). Also included
 CC are a host cell containing the nucleic acids (used to produce the
 CC proteins), the encoded proteins, an antibody that selectively binds to
 CC the polypeptide, and identifying a compound that binds to/modulates the
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
 CC methods are useful for diagnosing, treating cancer, aberrant cellular
 CC proliferation and/or differentiation, immune disorders, heart disorders,
 CC cardiovascular disorders including endothelial cell disorders,
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain
 CC and metabolic disorders, liver disorders and platelet disorders (many
 CC examples of these disorders are given in the specification). The present
 CC sequence is the Human glycoprotease encoded by cDNA 28472
 XX
 SQ Sequence 414 AA:
 Query Match 97.0%; Score 1203; DB 6; Length 414;
 Best Local Similarity 97.1%; Pred. No. 2.3e-126;
 Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 LLAIVQGVSDPFLLLKSLIDIAFGMDLVKVARLSLIKPECSWTSGGKAIETHAKQGRF 60
 DB 176 LLAIVQGVSDPFLLLKSLIDIAFGMDLVKVARLSLIKPECSWTSGGKAIETHAKQGRF 235
 QY 61 HPDIKPLHAKNCDPSFTGLQHTVDKTIIMKKKEGIEKKQIILSSADIAATVQHTMAC 120
 DB 236 HPDIKPLHAKNCDPSFTGLQHTVDKNNENKQEBEGIEKKQIILSSADIAATVQHTMAC 295

QY 121 HLVRTRRAILFCQKORDLLPQNNNAVLSVSGVSNFYIRRALEILTNAQTCTLLCPPPRL 180
 DB 296 HLVRTRRAILFCQKORDLLPQNNNAVLSVSGVSNFYIRRALEILTNAQTCTLLCPPPRL 355
 QY 181 CTDNGIMIAMNGIERLRAGLGIHDIESGIRYBPCPLGVDSIKVGEASIKVPLQKMEI 239
 DB 356 CTDNGIMIAMNGIERLRAGLGIHDIESGIRYBPCPLGVDSIKVGEASIKVPLQKMEI 414

RESULT 8

ADAS4471
 ID ADAS4471 standard; protein; 364 AA.

AC ADAS4471;

DT 20-NOV-2003 (first entry)

DE Human protein, SEQ ID 2039.

XX Cytosolic; Anti-inflammatory; Osteopathic; Neuroprotective; Nocotropic;

KM Gene Therapy; human; secretory protein; membrane proteins; cancer;

XX Inflammatory disease; osteoporosis; neurological disease.

XX Homo sapiens.

XX EPI293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.

XX 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isegai T, Shigiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX WPI; 2003-395539/38.

XX N-PSDB; ADA52832.

XX New polynucleotides encoding full-length polypeptides, e.g. secretory

XX and/or membrane proteins, useful for developing medicines for diseases in

XX which the gene is involved, or as target molecules for gene therapy.

XX Claim 14; SEQ ID NO 2039; 205PP; English.

XX The present invention relates to novel human secretory or membrane

XX proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-

XX ADA54071). The coding sequences are useful in the gene therapy of

XX diseases caused by abnormalities of the proteins, e.g. cancer,

XX inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 364 AA;

XX Query Match 77.4%; Score 960; DB 6; Length 364;

XX Best Local Similarity 98.4%; Pred. No. 5e-99;

XX Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 296 HLVRTRRAILFCQKORDLLPQNNNAVLSVSGVSNFYIRRALEILTNAQTCTLLCPPPRL 355
 QY 181 CTDNGIMIA 189
 DB 356 CTDNGIMIA 364

RESULT 9

ABB69133
 ID ABB69133 standard; protein; 409 AA.

AC ABB69133;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 34191.

XX Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL13236.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions.

XX Disclosure; SEQ ID NO 34191; 21PP + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

XX ABB72072). The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 409 AA;

XX Query Match 26.8%; Score 332.5; DB 4; Length 409;

XX Best Local Similarity 35.9%; Pred. No. 3.9e-28;

XX Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

XX 2 LALVGVSPFLILGSLDIPAGMDLKVARRSLIHPECSMTSGGKAIEHLAK-QGNRF 60
 DB 167 LVVANGPGRLTLGGTLDAPGBAPDKIRRRRLHLPEYRLWNGRAIEHAAQLSDPL 226
 QY 61 HFDIKRPLHAHAKNCFSTFGLOHTVDKTIKKKEKEGIEKGQILSSAADIATVQHTMAC 120
 DB 227 AYSFPLPLAQKQKCNCFSTFGITKNSFRRAIRAEARERTPPDGVISYGFCAILRSVSR 286
 QY 121 HLVRTRRAILFC--KORDLLPQNNNAVLSVSGVSNFYIRRALEILTNAQTCTLLCPPPRL 178
 DB 287 HLMHTTORAIRIERYCLPHRQLPQDPTPTLVMSGVSANNDIAIYNIHILAQYGRSFRPSK 346

Qy 179 RLCTDNGIMIAMNGIERLRAGILHDIIE-GIRYEPKPCGLVDISKVEGEA 228
Db 347 RYCSDNGMGMIAHGVGEOL-----LQDKRKASTRYDYD---SIDIGSAGFA 388

RESULT 10
ID AAY52216 standard; protein; 463 AA.
XX AAY52216;
XX AAY52216;
XX 09-FEB-2000 (first entry)
XX Arabidopsis thaliana yjld protein homologue.
XX
XX yjld protein; essential; Gram positive; Gram negative; conserved; motif;
XX identification; antagonist; antibacterial; antibiotic; broad spectrum;
XX treatment; infection; resistance; drug target.
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX Region 86..96
XX /note= "yjld conserved motif 3"
XX Region 111..131
XX /note= "yjld conserved motif 4"
XX Region 152..198
XX /note= "yjld conserved motif 2"
XX Region 208..259
XX /note= "yjld conserved motif 1"
XX
XX W0954470-A2.
XX
XX 28-OCT-1999.
XX
XX 20-APR-1999; 99WO-EP002635.
XX
XX 22-APR-1998; 98GB-00008423.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Arigoni F, Edgerton MD, Loferer H, Peltesch MC;
XX
XX WPI; 2000-013253/01.
XX
XX Novel bacterial polypeptides used to identify broad spectrum antibiotics.
XX
XX Claim 1; Fig 1; 55pp; English.
XX
XX Proteins AAY52202-Y52217 encompass a novel family of proteins designated
XX the yjld family, after the name given to the Escherichia coli family
XX member. These proteins are essential for the survival of both Gram
XX negative and Gram positive bacteria, although no function has as yet been
XX ascribed to these proteins. The yjld proteins, fragments of yjld proteins
XX (for example, fragments encompassing one or more conserved yjld motifs
XX such as AAY52218-Y52288) and nucleotides encoding them can be used to
XX identify antagonists and broad spectrum antibacterial compounds. These
XX antagonists and compounds can be used to treat a wide range of bacterial
XX infections. New antibiotics are urgently needed, as serious bacterial
XX infections and antibiotic resistant strains are becoming increasingly
XX prevalent. The proteins of the invention are essential proteins for
XX bacterial viability, and represent new targets for antibiotics
XX
XX Sequence 463 AA;
XX

Query Match 26.3%; Score 326; DB 3; Length 463;
Best Local Similarity 33.6%; Pred. No. 2.6e-27;
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

Qy 1 LILVGVGVSVFLLAGSLIDIAFMDLKVARRLSLIKPECSWSGRAIEHLAKQRRF 60
Db 221 LILVLAHLGCGYTOGLTIVDDAIGFAFDKATKMLGIDMH-----RSGGFAVERLEALGDAK 275

Qy 61 HPIKRPPLHAAKNCDSFSTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120
Db 276 SVKFNVPKMXCHDCNFSVAGLTKTQVRLALEAKE-----IRNRADIAASFQVAVL 325

Qy 121 HLVKRTHRAILFCRKQBDLLPQNNAVLVASGVAASNFYIRRALEITLNAQCTITLCPPL 180
Db 326 HLEEKCEKRAIDWALE---LEPSIKHNVISGVASNKYVRLNINVENKQLKLVCPPEPST 382

Qy 181 CTDNGIMIAMNGIERLRAGILHDIIEGIRYE-----PKCPLGVDISKX 224
Db 383 CTDNGVMVAMTGLERHFG-----RYDPPPTPEDEYVYDLRPRMPLGEEYAKG 432

Qy 225 VGEA 228
Db 433 RSEA 436

RESULT 11
AAG19287
ID AAG19287 standard; protein; 245 AA.
XX AAG19287;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21031.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX
XX 05-MAR-1999; 99US-0123180P.
XX
XX 09-MAR-1999; 99US-0123548P.
XX
XX 23-MAR-1999; 99US-0125788P.
XX
XX 25-MAR-1999; 99US-0125264P.
XX
XX 29-MAR-1999; 99US-0126785P.
XX
XX 01-APR-1999; 99US-0127462P.
XX
XX 06-APR-1999; 99US-0128234P.
XX
XX 08-APR-1999; 99US-0128714P.
XX
XX 16-APR-1999; 99US-0129845P.
XX
XX 19-APR-1999; 99US-0130077P.
XX
XX 21-APR-1999; 99US-0130449P.
XX
XX 23-APR-1999; 99US-0130510P.
XX
XX 23-APR-1999; 99US-0130891P.
XX
XX 28-APR-1999; 99US-0131449P.
XX
XX 30-APR-1999; 99US-0132048P.
XX
XX 30-APR-1999; 99US-0132407P.
XX
XX 04-MAY-1999; 99US-0132484P.
XX
XX 05-MAY-1999; 99US-0132485P.
XX
XX 06-MAY-1999; 99US-0132486P.
XX
XX 07-MAY-1999; 99US-0132487P.
XX
XX 11-MAY-1999; 99US-0132863P.
XX
XX 11-MAY-1999; 99US-0134256P.
XX
XX 14-MAY-1999; 99US-0134218P.
XX
XX 14-MAY-1999; 99US-0134219P.
XX
XX 14-MAY-1999; 99US-0134370P.
XX
XX 18-MAY-1999; 99US-0134768P.
XX
XX 19-MAY-1999; 99US-0134941P.
XX
XX 20-MAY-1999; 99US-0135124P.
XX
XX 21-MAY-1999; 99US-0135353P.
XX
XX 24-MAY-1999; 99US-0135629P.
XX
XX 25-MAY-1999; 99US-0136021P.
XX
XX 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0116782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142972P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.

PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150565P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0155486P.
PR 29-SEP-1999; 99US-0155596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160880P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157535P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160742P.
PR 21-OCT-1999; 99US-0160766P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 23-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 25.7%; Score 319; DB 3; Length 439;

Best Local Similarity 36.2%; Pred. No. 1,5e-26;

Matches 81; Conservative 31; Mismatches 82; Indels 30; Gaps 5;

QY 1 LIALVQVSDFLLKSLDIAPQMDIKYARRLSLIKHPBCSTMSGGKAI EHLAKQGNRF 60
DB 224 LTLVHLKQGYQLGTTVDIAIGEAFTAKYLGIDMH-----RSGGPAVEEHLALSGDAK 278
QY 61 HFDIKRPLHMANCPSPFGLQHVTDKLTIMKXKEKGEIKGQTLSSA-----ADIAA 112
DB 279 SVKFNVPKXKRDKNFSGYGLTKTVRLALEAKE---IDAKCPVSSATYEDRRNRADIAA 334
QY 113 TVQHTWACHLVKRTIRALIFCKORDLLPQNNAVLVASGVASNFYIRRALEITLNATQCT 172
DB 335 SFQRAVVLHBEKRCERATDMALE---LEPSIKHMTVSGGVASNKYRRLNNIIVENKNTLK 391
QY 173 LLCPPRLCTDNGIMTANNGIERLRAGLGLHDIGIRYEPKCP 216
DB 392 LVCPPSLCTDNGVVMVAMTGLEHFRVG-----RYDPPPP 425

RESULT 13

AAAG19285
ID AAG19285 standard; protein; 444 AA.

XX AAG19285;
AC

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 21029.

DB Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000BP-00301439.

PF 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134212P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142920P.
PR 13-JUL-1999; 99US-0142977P.
PR 14-JUL-1999; 99US-0143543P.
PR 15-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144633P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.

PR 26-JUL-1999; 99US-0145376P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147432P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150848P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.

PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161358P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161932P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 25.7%; Score 319; DB 3; Length 444;
 Best Local Similarity 36.2%; Pred. No. 1.5e-26;

Matches 81; Conservative 31; Mismatches 82; Indels 30; Gaps 5;

QY 1 LIALVGVSDPFLIGKSLDIPAGMDLVKVARRLSLIKHPCSTMSGGAIEHLAKOGRNF 60
 DB 229 LVLVLAHKLGQYTOGTVDVAIGBAFDYAKWLGIDMH-----RSGGPAVEBLALEGPAK 283
 QY 61 HFDIKPPLHAKNCDPSFTGLQHVTDKILMKKEKEGIEKGQILSSA-----ADTAA 112
 DB 284 SVKKNVPEKTKHKKDCNFSYAGLKTQVRLAIEAKE-----IDACKPVSSATNEDRRNRADTAA 339
 QY 113 TVQHTMACHLVKRTHRALIFCKQRDLPQNNAVLVASGVASNFYIRRALBITNATQCT 172
 DB 340 SFQGVAVLHEKEKERALDWMAL---LEPSIKHWVISGGSVANSKTVRLANNTVENKMLK 396
 QY 173 LACPPRLCTDNGIMIAMNGIERLPAIGILHIDIGIRYBPKCP 216
 DB 397 LVCPPPSLCTDNGVAVWMTGLEHFRVG-----RYDPPP 430

RESULT 14

ABU22934
 ID ABU22934 standard; protein; 382 AA.

XX AC ABU22934;

XX DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #8461.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bordetella pertussis.

PN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.

XX N-PSDB; ACA26804.

XX PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 50858; 1766bp; English.
 CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-regulated gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 382 AA;

Query Match 20.7%; Score 257; DB 6; Length 382;

Best Local Similarity 36.1%; Pred. No. 1.2e-19;

Matches 74; Conservative 33; Mismatches 78; Indels 20; Gaps 7;

QY 2 LIALVGVSDPFLIGKSLDIPAGMDLVKVARRLSLIKHPCSTMSGGAIEHLAKOGRNF 61
 DB 178 LMLVGVGVEYELLSGTLDDPAAGEAFDPSKAKMLGL-GYP-----GGPALARLAEQDDASR 230
 QY 62 PDIKPPLHAKNCDPSFTGLQ-HVTDKILMKKEKEGIEKGQILSSAADIATVQHTMAC 120
 DB 231 YDLPRPMLHSGDLDSFSGLKTAVLTV--KATADGSELGR--QDRADLAATQAAIYB 286
 QY 121 HLVKRTHRALIFCKQRDLPQNNAVLVASGVASNFYIRRALBITNATQCTLLCPPRL 180
 DB 287 VLAARAIRAL---KQTGL-----RLTVVAGVGAVNALLRAHLARALKPRAEAYFPPLSL 338
 QY 181 CTDNGIMIAMNGIERLPAIGILHID 205
 DB 339 CTDNGAMIAFAAAERVKAGLADIRE 363

RESULT 15

ABU35232
 ID ABU35232 standard; protein; 348 AA.

XX AC ABU35232;

XX DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #20759.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Moraxella catarrhalis.

PN WO200277183-A2.

PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.

XX N-PSDB; ACA39102.

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 63156; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway of
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 348 AA:

Query Match 20.0%; Score 248.5; DB 6; Length 348;

Best Local Similarity 33.2%; Pred. No. 9.4e-19;
Matches 66; Conservative 33; Mismatches 79; Indels 21; Gaps 5;

QY 1 LIALVGVSDPLLGKSLDIPRGMDLVARRLSLIKHPCSTMSGKAIEHLAKQGRF 60
DB 147 MLVRADGCVGIQIGESTIDVAVGCPDTAKLKL-PPY-----GGPIREKDAKNQNP 199
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKIMKKKEKEGIBKQIUSADIAATVQHTAC 120
DB 200 AYELPRMQR-KGLDPFSRSGKTAIHNIKDTPNAGSDP-----ATRADIASPEYAVVD 253
QY 121 HLVRTRRAILFCQKQDILLPQNNAVLVASGVASNFYIRALBILITNAQTCLCPPL 180
DB 254 TLVAKCTKALQMTGIRKQ-----LVVAGGVASANQMLRRTLTETLRQIDASVYAPTEL 305
QY 181 CTDNGIMIAMNGIRRLRAG 199
||||| |||: | | |

DB 306 CTDNGAMIAYAGFCRLSCG 324

Search completed: February 16, 2005, 13:05:57
Job time : 49.0598 secs

GenCore version 5.1.6 -
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:43:41 ; Search time 14.288 Seconds
(without alignments)
1248.675 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVQGVSDPFLILGKSLDI.....DISKEVGEASIKVPLKMEI 239

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/6D.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	267	US-10-067-443-22	Sequence 22, Appl
2	1240	100.0	414	US-10-067-443-2	Sequence 2, Appl
3	1213.5	97.9	439	US-10-067-443-19	Sequence 19, Appl
4	326	26.3	463	US-10-067-443-3	Sequence 3, Appl
5	247	19.9	350	US-09-540-236-2726	Sequence 2726, Ap
6	231	18.6	401	US-09-252-991A-17372	Sequence 17372, A
7	212	17.1	357	US-09-543-681A-6513	Sequence 6513, Ap
8	209.5	16.9	342	US-08-087-797-3	Sequence 3, Appl
9	205	16.5	343	US-09-489-039A-9221	Sequence 9221, Ap
10	201.5	16.2	421	US-10-067-443-4	Sequence 4, Appl
11	201.5	16.2	421	US-10-067-443-28	Sequence 28, Appl
12	200	16.1	325	US-08-087-797-2	Sequence 2, Appl
13	197	15.9	363	US-09-107-532A-6609	Sequence 6609, Ap
14	193.5	15.6	327	US-09-107-532A-6609	Sequence 5, Appl
15	190	15.3	336	US-08-987-121A-4	Sequence 4, Appl
16	189	15.2	273	US-09-710-279-728	Sequence 728, Appl
17	187	15.1	344	US-09-602-777A-148	Sequence 148, Appl
18	186	15.0	335	US-08-961-083-52	Sequence 52, Appl
19	186	15.0	335	US-09-536-784-52	Sequence 52, Appl
20	186	15.0	368	US-09-134-001C-3909	Sequence 3909, Ap
21	184	14.8	336	US-09-066-512-2	Sequence 2, Appl
22	182	14.7	336	US-09-583-110-4857	Sequence 4857, Ap
23	181.5	14.6	336	US-09-328-352-4387	Sequence 4387, Ap
24	177	14.3	326	US-09-107-433-4221	Sequence 4221, Ap
25	174.5	14.1	344	US-09-198-452A-213	Sequence 213, Appl
26	174.5	14.1	360	US-09-438-185A-196	Sequence 196, Appl
27	174	14.0	341	US-09-149-624-2	Sequence 2, Appl

28	173	14.0	366	US-09-134-000C-4956	Sequence 4956, Ap
29	148.5	12.0	340	US-10-067-443-6	Sequence 6, Appl
30	117	9.4	292	US-09-724-623-81	Sequence 81, Appl
31	94.5	7.6	480	US-09-583-110-5050	Sequence 5050, Ap
32	94.5	7.6	481	US-09-107-433-3197	Sequence 3197, Ap
33	89.5	7.2	214	US-09-328-352-4609	Sequence 4609, Ap
34	83	6.7	42	US-09-902-540-13841	Sequence 13841, A
35	81.5	6.5	1087	US-08-264-002-5	Sequence 5, Appl
36	81	6.5	1072	US-09-902-540-15572	Sequence 15572, A
37	79	6.4	445	US-08-083-945C-2	Sequence 2, Appl
38	79	6.4	445	US-08-083-945C-7	Sequence 7, Appl
39	79	6.4	578	US-09-538-092-968	Sequence 968, Appl
40	78	6.3	334	US-08-530-950-4	Sequence 4, Appl
41	78	6.3	334	US-08-530-950-4	Sequence 4, Appl
42	78	6.3	334	US-08-576-240-2	Sequence 2, Appl
43	78	6.3	334	US-08-888-429A-4	Sequence 4, Appl
44	78	6.3	334	US-09-149-879-4	Sequence 4, Appl
45	78	6.3	334	US-09-057-009-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-10-067-443-22
; Sequence 22, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-22
Query Match 100.0%; Score 1240; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 5.6e+145;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LIALVQGVSDPFLILGKSLDIAPGDMIDKVARRLSLIKPECSMTSGGKAIIEHLAKQGNRP 60
DB 29 LIALVQGVSDPFLILGKSLDIAPGDMIDKVARRLSLIKPECSMTSGGKAIIEHLAKQGNRF 88
QY 61 HEDIPPLHMANCPSPFGIOHVTDKITMKKEKEGIEKGOILSADIAITVQHTMAC 120
DB 89 HEDIPPLHMANCPSPFGIOHVTDKITMKKEKEGIEKGOILSADIAITVQHTMAC 148
QY 121 HLKERTHALLFCCKRDLLPNNNAVTVASGVASNFYIRALIEILTNATQCTLLCPPL 180
DB 149 HLKERTHALLFCCKRDLLPNNNAVTVASGVASNFYIRALIEILTNATQCTLLCPPL 208
QY 181 CTGNGIMIAMNGIEERLRAIGILHDIIEGIRYEPKCELVADISKEVGEASIKVPOLKMEI 239
DB 209 CTGNGIMIAMNGIEERLRAIGILHDIIEGIRYEPKCELVADISKEVGEASIKVPOLKMEI 267
RESULT 2
US-10-067-443-2
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company

1 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
2 FILE REFERENCE: D0073 NP
3 CURRENT APPLICATION NUMBER: US/10/067,443
4 PRIOR FILING DATE: 2002-02-05
5 PRIOR APPLICATION NUMBER: US 60/266,518
6 PRIOR FILING DATE: 2001-02-05
7 PRIOR APPLICATION NUMBER: US 60/282,814
8 PRIOR FILING DATE: 2001-04-10
9 NUMBER OF SEQ ID NOS: 71
10 SOFTWARE: PatentIn version 3.0
11 SEQ ID NO 2
12 LENGTH: 414
13 TYPE: PRT
14 ORGANISM: Homo sapiens
US-10-067-443-2

Query Match 100.0%; Score 1240; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 1,1e-144;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAIVGVSDFLILGKSLDIPGMDLVARRLSLIKPECSMTSGKAIEHLAKQNR 60
DB 176 LLAIVGVSDFLILGKSLDIPGMDLVARRLSLIKPECSMTSGKAIEHLAKQNR 235
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120
DB 236 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 295
QY 121 HLKRTTRAILFCQKQDILLPNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 180
DB 296 HLKRTTRAILFCQKQDILLPNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 355
QY 181 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKCPPLGVDISKVEASIKVPOLKMEI 239
DB 356 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKCPPLGVDISKVEASIKVPOLKMEI 414

RESULT 3
US-10-067-443-19
1 Sequence 19, Application US/10067443
2 Patent No. 6642041
3 GENERAL INFORMATION:
4 APPLICANT: Bristol-Myers Squibb Company
5 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
6 FILE REFERENCE: D0073 NP
7 CURRENT APPLICATION NUMBER: US/10/067,443
8 PRIOR FILING DATE: 2002-02-05
9 PRIOR APPLICATION NUMBER: US 60/266,518
10 PRIOR FILING DATE: 2001-02-05
11 PRIOR APPLICATION NUMBER: US 60/282,814
12 PRIOR FILING DATE: 2001-04-10
13 NUMBER OF SEQ ID NOS: 71
14 SOFTWARE: PatentIn version 3.0
15 SEQ ID NO 19
16 LENGTH: 439
17 TYPE: PRT
18 ORGANISM: homo sapiens
US-10-067-443-19

Query Match 97.9%; Score 1213.5; DB 4; Length 439;
Best Local Similarity 90.2%; Pred. No. 2.3e-141;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 LLAIVGVSDFLILGKSLDIPGMDLVARRLSLIKPECSMTSGKAIEHLAKQNR 60
DB 176 LLAIVGVSDFLILGKSLDIPGMDLVARRLSLIKPECSMTSGKAIEHLAKQNR 235
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120
DB 236 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 295

QY 99 ---EKQILSSADIAATVQHTMACHLVRRTRAILFCQKQDILLPNNAVLVASGVASN 155
DB 296 CPEKQILSSADIAATVQHTMACHLVRRTRAILFCQKQDILLPNNAVLVASGVASN 355
QY 156 FYIRALBITLNAQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPK 215
DB 356 FYIRALBITLNAQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPK 415
QY 216 PLGVDSKEVGRASIKVPOLKMEI 239
DB 416 PLGVDSKEVGRASIKVPOLKMEI 439

RESULT 4
US-10-067-443-3
1 Sequence 3, Application US/10067443
2 Patent No. 6642041
3 GENERAL INFORMATION:
4 APPLICANT: Bristol-Myers Squibb Company
5 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
6 FILE REFERENCE: D0073 NP
7 CURRENT APPLICATION NUMBER: US/10/067,443
8 PRIOR FILING DATE: 2002-02-05
9 PRIOR APPLICATION NUMBER: US 60/266,518
10 PRIOR FILING DATE: 2001-02-05
11 PRIOR APPLICATION NUMBER: US 60/282,814
12 PRIOR FILING DATE: 2001-04-10
13 NUMBER OF SEQ ID NOS: 71
14 SOFTWARE: PatentIn version 3.0
15 SEQ ID NO 3
16 LENGTH: 463
17 TYPE: PRT
18 ORGANISM: Arabidopsis thaliana
US-10-067-443-3

Query Match 26.3%; Score 326; DB 4; Length 463;
Best Local Similarity 33.6%; Pred. No. 2.7e-31;
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 1 LLAIVGVSDFLILGKSLDIPGMDLVARRLSLIKPECSMTSGKAIEHLAKQNR 60
DB 221 LLAIVGVSDFLILGKSLDIPGMDLVARRLSLIKPECSMTSGKAIEHLAKQNR 275
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120
DB 276 SVKFNVPKYNKDCNFSYAGLQVRLAIEAKE-----INRRADIAASFQVAVL 325
QY 121 HLKRTTRAILFCQKQDILLPNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 180
DB 326 HLBKCEKRAIDWALE---LEPSIKHNVISGVASNRYRLNINVENKRLKLVCPPL 382
QY 181 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKCPPLGVDISKVEASIKVPOLKMEI 239
DB 383 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKCPPLGVDISKVEASIKVPOLKMEI 432
QY 225 VGEA 228
DB 433 RSEA 436

RESULT 5
US-09-540-236-2726
1 Sequence 2726, Application US/09540236
2 Patent No. 6673910
3 GENERAL INFORMATION:
4 APPLICANT: Gary L. Breton et al.
5 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
6 FILE REFERENCE: 2709.2005-001
7 CURRENT APPLICATION NUMBER: US/09/540,236
8 PRIOR FILING DATE: 2000-04-04
9 NUMBER OF SEQ ID NOS: 3840

SEQ ID NO 2726
LENGTH: 350
TYPE: PRT
ORGANISM: M. catarrhalis
US-09-540-236-2726

Query Match 19.9%; Score 247; DB 4; Length 350;
Best Local Similarity 31.5%; Pred. No. 1,1e-21;
Matches 70; Conservative 34; Mismatches 80; Indels 38; Gaps 6;

1 LALVGVSDPFLILGKSLDIAPGMDLVKARSLIKHEPCSTMSGKAIIEHLAKGNRF 60
149 MLVADGVGVQIIGESIDDAVGEFCFDTAKLKL-PYP-----GCPNIEKTLAKGNPH 201
61 HFDIKPRLHAKNCDFSTGQHTVDKTIIMKKEEGIEKQIILSSADIAATVQHTMAC 120
202 AYELPRPQOR-KGLDFSPSGMKTAIHNLIKOTPNASQSP-----ATRADIASFEYAVVD 255
121 HLKRTTRAILFCQKRDLLPQNNAVLVASGVASNFYIRALEITLNATQCTLLCPPRL 180
256 TLVKKCKKALQMTGIRQ-----LVVAGVSANQTLRRLTETLRQIDASVYVAPREL 307
181 CTDNGIMIAMNGIERLRAG-----IGILHD 205
308 CTDNGAMIAVAGFCRLSGQSDILA VRCTPRWDMTLGIERYD 349

RESULT 6
US-09-252-991A-17372
Sequence 17372, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17372
LENGTH: 401
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17372

Query Match 18.6%; Score 231; DB 4; Length 401;
Best Local Similarity 33.8%; Pred. No. 1,3e-19;
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

2 LALVGVSDPFLILGKSLDIAPGMDLVKARSLIKHEPCSTMSGKAIIEHLAKGNRF 61
202 LVRVDGIGRQQLGSDVDDAGAPDKAKLIGL-GYP-----GCPRIARLAERGTRGR 254
62 FDIKPLHAKNCDFSTGQHTVDKTIIMKKEEGIEKQIILSSADIAATVQHTMAC 121
255 FVPRPMTDRPGLDFSGKLTFTLN-TWQRCVEAGDSDSEQ---TRCIDALAFQAVVET 310
122 LVKRTTRAILFCQKRDLLPQNNAVLVASGVASNFYIRALEITLNATQCTLLCPPRL 181
311 LLIKCRBAL---KQGTG-L-KN---LVLAGVSANQTLRRLTETLRQIDASVYVAPREL 362
182 TDNGIMIAMNGIERLRAGLIGILHDIEGIRYEPKCP 217
363 TDNGAMIAVAGFCRLSGQSDILA VRCTPRWDMTLGIERYD 395

RESULT 7
US-09-543-681A-6513
Sequence 6513, Application US/09543681A

Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 6513
LENGTH: 357
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-6513

Query Match 17.1%; Score 212; DB 4; Length 357;
Best Local Similarity 30.3%; Pred. No. 2,4e-17;
Matches 66; Conservative 32; Mismatches 90; Indels 30; Gaps 7;

2 LALVGVSDPFLILGKSLDIAPGMDLVKARSLIKHEPCSTMSGKAIIEHLAKGNRF 61
159 LLSVIGIEBYTLIGSIDDAAGEAPDKAKLIGL-DYP-----GCPVLKMAQGVGGR 211
62 FDIKPLHAKNCDFSTGQHTVDKTIIMKKEEGIEKQIILSSADIAATVQHTMAC 121
212 FVPRPMTDRPGLDFSGKLTFTLN-TWQRCVEAGDSDSEQ---TRCIDALAFQAVVET 263
122 LVKRTTRAILFCQKRDLLPQNNAVLVASGVASNFYIRALEITLNATQCTLLCPPRL 180
264 LAIKRRA-----LEQTFKRLVAGVSANQTLRRLTETLRQIDASVYVAPREL 314
181 CTDNGIMIAMNGIERLRAGLIGILHDIEGIRYEPKCP 217
315 CTDNGAMIAVAGFCRLSGQSDILA VRCTPRWDMTLGIERYD 349

RESULT 8
US-08-087-797-3

Sequence 3, Application US/08087797
Patent No. 5543312
GENERAL INFORMATION:

APPLICANT: Mellors, Alan
APPLICANT: Lo, Reggie Y.C.
APPLICANT: Abdullah, Khalid M.
TITLE OF INVENTION: Pasteurella Haemolytica
TITLE OF INVENTION: Glycoprotease
TITLE OF INVENTION: Gene and the Purified Enzyme
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Bell, Seltzer, Park & Gibson, P.A.
STREET: 1211 East Morehead Street,
CITY: Charlotte
STATE: No. 5543312th Carolina
COUNTRY: United States
ZIP: 28234

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, version #1.25
APPLICATION DATA:
APPLICATION NUMBER: US/08/087,797
FILING DATE: 14-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Layton, Jr., Samuel G.
REGISTRATION NUMBER: 22907
REFERENCE/DOCKET NUMBER: 3374-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704 377 1561
TELEFAX: 704 334 2014

/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 342 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ US-08-087-797-3

Query Match 16.9%; Score 209.5; DB 1; Length 342;
Best Local Similarity 30.9%; Pred. No. 4.7e-17;
Matches 67; Conservative 30; Mismatches 97; Indels 23; Gaps 7;

QY 2 LALVQGVSDPFLILGKSLDIAPGDMLDKVARLSLIKHEPCSTMSGKAIIEHLAKGNFH 61
DB 142 LISTVGIQYELLESIDDAAGEAFDKTAKLGL-DYP-----GGPLISKPAAGSTAGR 194
QY 62 FDIKPLHAKNCFSTFGLOHVTDKIIMKKEKEGIEKGQILSSAADIATVQHTMACH 121
DB 195 FVPRPMTDRPGDLPSFGSLKTPAANTIRDNXXXXGXDDQ---TRADIAAFEDAVYDT 251
QY 122 LVKRTTRAILFCQKQDILLPÖNN-AVLVASGVASNFYIRRALEILTNAQCTLLCPPL 180
DB 252 LMICKKRA-----LDQTFKRLVMAGVGSANRTIRAKIAEMKKRGSEVFYARPEF 302
QY 181 CTONGIMAMNGIERLRAGLGIHIDIEGIRYEPKPL 217
DB 303 CTONGAMIAVAGVRLQFGA---GATADL-GVSVPRPWL 336

RESULT 9

US-09-489-039A-9221
/ Sequence 9221, Application US/09489039A
/ Patent No. 6610836
/ GENERAL INFORMATION:
/ APPLICANT: Gary Breton et. al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/ TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 2709.2004001
/ CURRENT APPLICATION NUMBER: US/09/489,039A
/ PRIOR FILING DATE: 2000-01-27
/ PRIOR APPLICATION NUMBER: US 60/117,747
/ PRIOR FILING DATE: 1999-01-29
/ NUMBER OF SEQ ID NOS: 14342
/ SEQ ID NO 9221
/ LENGTH: 343
/ TYPE: PRT
/ ORGANISM: Klebsiella pneumoniae
/ US-09-489-039A-9221

Query Match 16.5%; Score 205; DB 4; Length 343;
Best Local Similarity 30.3%; Pred. No. 1.7e-16;
Matches 66; Conservative 29; Mismatches 93; Indels 30; Gaps 7;

QY 2 LALVQGVSDPFLILGKSLDIAPGDMLDKVARLSLIKHEPCSTMSGKAIIEHLAKGNFH 61
DB 148 LISTVGIQYELLESIDDAAGEAFDKTAKLGL-DYP-----GGPLISKPAAGSTAGR 200
QY 62 FDIKPLHAKNCFSTFGLOHVTDKIIMKKEKEGIEKGQILSSAADIATVQHTMACH 121
DB 201 FVPRPMTDRPGDLPSFGSLKTPAANTIRSNDD-----QTRADIAAFEDAVYDT 252
QY 122 LVKRTTRAILFCQKQDILLPÖNN-AVLVASGVASNFYIRRALEILTNAQCTLLCPPL 180
DB 253 LMICKKRA-----LEQTFKRLVMAGVGSANRTIRAKIAEMKKRGSEVFYARPEF 303
QY 181 CTONGIMAMNGIERLRAGLGIHIDIE-GIRYEPKPL 217
DB 304 CTONGAMIAVAGVRLQFGA---KAEIGVTVPRPWL 337

RESULT 10
US-10-067-443-4

/ Sequence 4, Application US/10067443
/ Patent No. 6642041
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
/ FILE REFERENCE: D0073 NP
/ CURRENT APPLICATION NUMBER: US/10/067,443
/ PRIOR FILING DATE: 2002-02-05
/ PRIOR APPLICATION NUMBER: US 60/266,518
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: US 60/282,814
/ PRIOR FILING DATE: 2001-04-10
/ NUMBER OF SEQ ID NOS: 71
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 4
/ LENGTH: 421
/ TYPE: PRT
/ ORGANISM: Caenorhabditis elegans
/ US-10-067-443-4

Query Match 16.2%; Score 201.5; DB 4; Length 421;
Best Local Similarity 27.5%; Pred. No. 6.3e-16;
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

QY 1 LALVQGVSDPFLILGKSLDIAPGDMLDKVARLSLIKHEPCSTMSGKAIIEHLAKGN-- 58
DB 163 LTVSADVEKFLVYQSGSPBCIDKVARQDGL-GSEFGGIGHVAIVELIASPASD 221
QY 59 -RFHFDIKPLHAKNCFSTFGLOHVTDKIIMKKEKEGIEKGQILSSAADIATVQHT 117
DB 222 GHLRYPILPENVPKANMNFQDKGSTANLELRKNSFISID-----IPFGASLQNT 274
QY 118 MACLVKRTTRAILFCQKQDILLPÖNNAVLVASGVASNFYIRRALEILTNAQCTLLCP 177
DB 275 VARHISSKLHIFPESISEQEKLPKQ---LVIGGVANQYIRGALSLSAAHNVITIKYL 331
QY 178 PRLCTDNGIMAMNGIERLRAGLGIHIDIEGIRYEP-----KCPLGVDISKEVG 226
DB 332 LSLCTDNAMEIAVSGL-----LMLVNRSEALIMWRNDIPDTYAHARSDIGDASSEI- 384
QY 227 EASIKVPOLKM 237
DB 365 ---IDTPRRKL 392

RESULT 11
US-10-067-443-28
/ Sequence 28, Application US/10067443
/ Patent No. 6642041
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
/ FILE REFERENCE: D0073 NP
/ CURRENT APPLICATION NUMBER: US/10/067,443
/ PRIOR FILING DATE: 2002-02-05
/ PRIOR APPLICATION NUMBER: US 60/266,518
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: US 60/282,814
/ PRIOR FILING DATE: 2001-04-10
/ NUMBER OF SEQ ID NOS: 71
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 28
/ LENGTH: 421
/ TYPE: PRT
/ ORGANISM: Caenorhabditis elegans
/ US-10-067-443-28

Query Match 16.2%; Score 201.5; DB 4; Length 421;
Best Local Similarity 27.5%; Pred. No. 6.3e-16;
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

QY 1 LIALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWMSGKAIEHLAKQGN-- 58
Db 163 LISVAEVEKFKLYGVSSGSGPCIDKVARQLDGL-GSEFGIHVGAIVELASRASAD 221
QY 59 -RHFEDIKPPLHAKNCDPFTGLQHTDKIIMKEKEEGIEKGQILSSADIAATVQHT 117
Db 222 GHLRYPIFLPVPKANNFDDIKSGSYLNLRLRKNSETSID-----IPDFCASLQNT 274
QY 118 MACHLVKTRRALIFCQKORDLPQNNAVLVASGVASNFYIRBALEILTNATQCTLLCP 177
Db 275 VARIHSSKLIHFFESLSEQEDLPKQ--LVIGGVANQYIFGALSKLSAANNVTTIKVL 331
QY 178 PRLCTDNGIMIANNGIERLRAGLILHDIEGIRYEP-----KCPGLVDISKVWG 226
Db 332 LSLCTDAEMHAYSGL-----LMLVNRSEALIMWRPNIPDTYAHASDICTDASSSEI- 384
QY 227 EASIKVPOLQM 237
Db 385 ---IDTPRRKL 392
RESULT 12
US-08-087-797-2
; Sequence 2, Application US/08087797
; Patent No. 5543312
; GENERAL INFORMATION:
; APPLICANT: Mellore, Alan
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Abdullah, Khalid M.
; TITLE OF INVENTION: Pasteurella Haemolytica
; TITLE OF INVENTION: Glycoprotease
; TITLE OF INVENTION: Gene and the Purified Enzyme
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street,
; CITY: Charlotte
; STATE: NC 5543312th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,797
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Layton, Jr., Samuel G.
; REGISTRATION NUMBER: 22807
; REFERENCE/DOCKET NUMBER: 3374-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704 377 1561
; TELEFAX: 704 334 2014
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-087-797-2
Query Match 16.1%; Score 200; DB 1; Length 325;
Best Local Similarity 31.6%; Pred. No. 6.5e-16;
Matches 62; Conservative 25; Mismatches 91; Indels 18; Gaps 5;
QY 2 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWMSGKAIEHLAKQGNFPH 61
Db 142 LVVNDGVGQYBILGESIDDAAGEAFDKTKLGL-DY-----AGVMSKLAEGSTNR 194
QY 62 FDIKPLHAKNCDPFTGLQHTDKIIMKEKEEGIEKGQILSSADIAATVQHTMAC 121

Db 195 FEPFRMTDRPGIDFSFSLKTFEANTIKANINENELDEQ--TKCDIAHAFQAV-- 248
QY 122 LKRTTRALIFCQKORDLPQNNAVLVASGVASNFYIRBALEILTNATQCTLLCP 181
Db 249 ---VDTLIKK-RALDEGTGYKRLVMAGVANKQLRADLAEWMKKLKEVEYFPPRQFC 303
QY 182 TDNGIMIANNGIERLR 197
Db 304 TDNGAMIAATGFLRLK 319
RESULT 13
US-09-107-532A-6609
; Sequence 6609, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Denek
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6609:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...363
; SEQUENCE DESCRIPTION: SEQ ID NO: 6609:
US-09-107-532A-6609
Query Match 15.9%; Score 197; DB 4; Length 363;
Best Local Similarity 32.6%; Pred. No. 1.8e-15;
Matches 63; Conservative 34; Mismatches 64; Indels 32; Gaps 9;
QY 2 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWMSGKAIEHLAKQGNF 60
Db 169 LVVNDGVGSEYBILGETRBDAAAGEAYDKGVVGL-----SYPSGKBIIDQLAHQCKDNY 221
QY 61 FDIKPLHAKNCDPFTGLQHTDKIIMKEKEEGIEKGQILSSADIAATVQHTMAC 119

Db 222 HF--PRAMIHEDNDYDFSGLSKSAFINLVHNAOGRGDLDRN-----DLAASFOASVI 272
QY 120 CHVKTGTRAILFCQKORDLPPONNAV--LVASGVASNFYIRRALBITLNA--TQCTLLC 175
Db 273 DVLTKTLRA---C-----QNPVKQLVAVGVAAVNOGLBGLQALASLAPLPEVELVI 322
QY 176 PPPRLCTDNGIMI 188
Db 323 PPLRLCGDNAMMI 335

RESULT 14

US-10-067-443-5
; Sequence 5, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-10-067-443-5

Query Match 15.6%; Score 193.5; DB 4; Length 327;
Best Local Similarity 28.9%; Pred. No. 4.2e-15;
Matches 55; Conservative 38; Mismatches 68; Indels 29; Gaps 6;

QY 13 LKGSLDIAPGMDLKVARRSLIKHPECSTMSGKALEHLAKQNRHFDPKPLHNAK 72
Db 151 VLGETLDSAGEADPKVARLIGL-GYP-----GGPVIDRAVKGDPEKYSFPRMLDD 203
QY 73 NCDFSFTGLQHTDKIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKTRAILF 132
Db 204 SYNSFAGLK--TSVLYFLQREKGYK-----VEDVAASQKAVDILVEKTR---- 249
QY 133 CKORDLPPONNAV--LVASGVASNFYIRRALBITLNACTLLCPPRLCTDNGIMIAM 190
Db 250 -----LARNIGIRKIAFVGVAAVNSMLREYVRKABRMNVEVFPPLICTDNALMVAK 303
QY 191 NGIERLPAQL 200
Db 304 AGYEKAKRGM 313

RESULT 15

US-08-987-121A-4
; Sequence 4, Application US/08987121A
; GENERAL INFORMATION:
; APPLICANT: Hoskins, Jo Ann
; APPLICANT: Tang, Joseph Chio-Chung
; APPLICANT: Treadway, Patil Jean
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/987,121A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-987-121A-4

Query Match 15.3%; Score 190; DB 3; Length 336;
Best Local Similarity 31.1%; Pred. No. 1.2e-14;
Matches 61; Conservative 32; Mismatches 73; Indels 30; Gaps 8;

QY 2 LALVGVSDPFLIGKSLDIAPGMDLKVARRSLIKHPECSTMSGKALEHLAKQNRHF 61
Db 144 LVVSBAGDYKIVGETRDVAGEAYDKGVNGL-----TYPAGREIDELAHQCHDI- 195
QY 62 FDIKPELHNAKNCDFSFTGLQHTDKIIMKKEKEGIEKQILSSAADIATVQHTMACH 121
Db 196 YDPPRAMIKEDYLBERSFSGLSKSAFINHNAE-----QNGESLST-EDLCASQAANVMI 249
QY 122 LVKTRAILFCQKORDLPPONNAVIVASGVASNFYIRRALBITLNACTLLCPPRLC 181
Db 250 LMAKTKKAL-----EKYPVK--TLVAVAGVAAVANGLRERL--ATRTIVNVVILPRLIC 299
QY 182 TDNGIMIA-----WN 191
Db 300 GDNAGMIAYASVSEWN 315

Search completed: February 16, 2005, 13:10:26
Job time : 15.288 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:08:47 ; Search time 138.464 Seconds
(without alignments)
563.995 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVQGVSDFLILGKSLDI.....DISKVEGASIKVPLKMEI 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptcodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptcodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptcodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptcodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptcodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptcodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptcodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptcodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptcodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptcodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptcodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptcodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptcodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptcodata/2/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptcodata/2/pubpaa/US10D_PUBCOMB.pep:*
16: /cgn2_6/ptcodata/2/pubpaa/US10E_PUBCOMB.pep:*
17: /cgn2_6/ptcodata/2/pubpaa/US10F_PUBCOMB.pep:*
18: /cgn2_6/ptcodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptcodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptcodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	267	14	US-10-067-443-22 Sequence 22, Appl
2	1240	100.0	267	15	US-10-649-273-22 Sequence 22, Appl
3	1240	100.0	267	15	US-10-651-722-22 Sequence 22, Appl
4	1240	100.0	414	14	US-10-067-443-22 Sequence 2, Appl
5	1240	100.0	414	15	US-10-649-273-22 Sequence 2, Appl
6	1240	100.0	414	15	US-10-651-722-22 Sequence 2, Appl
7	1213.5	97.9	439	14	US-10-067-443-19 Sequence 19, Appl
8	1213.5	97.9	439	15	US-10-649-273-19 Sequence 19, Appl
9	1213.5	97.9	439	15	US-10-651-722-19 Sequence 19, Appl
10	1202	97.0	414	14	US-10-012-140-5 Sequence 5, Appl
11	960	77.4	364	14	US-10-094-749-2039 Sequence 2039, Ap
12	337	27.2	445	15	US-10-424-599-209259 Sequence 209259, Ap
13	326	26.3	463	14	US-10-067-443-3 Sequence 3, Appl

14	326	26.3	463	15	US-10-649-273-3	Sequence 3, Appl
15	326	26.3	463	15	US-10-651-722-3	Sequence 3, Appl
16	270	21.8	444	15	US-10-437-963-113732	Sequence 113732, A
17	257	20.7	382	15	US-10-282-122A-50858	Sequence 50858, A
18	254	20.5	333	14	US-10-012-140-25	Sequence 25, Appl
19	248.5	20.0	338	15	US-10-282-122A-63156	Sequence 63156, A
20	247	19.9	312	15	US-10-282-122A-44499	Sequence 44499, A
21	241	19.4	343	15	US-10-282-122A-67227	Sequence 67227, A
22	239	19.3	251	14	US-10-081-051-9	Sequence 9, Appl
23	231	18.6	341	9	US-09-815-242-11798	Sequence 11798, A
24	231	18.6	341	15	US-10-282-122A-66200	Sequence 66200, A
25	228.5	18.4	347	14	US-10-012-140-24	Sequence 24, Appl
26	224	18.1	342	9	US-09-815-242-11043	Sequence 11043, A
27	224	18.1	342	15	US-10-282-122A-58204	Sequence 58204, A
28	213	17.2	337	15	US-09-815-242-13780	Sequence 13780, A
29	213	17.2	337	15	US-10-282-122A-75485	Sequence 75485, A
30	212	17.1	340	14	US-10-282-122A-68438	Sequence 68438, A
31	211	17.0	41	14	US-10-012-140-26	Sequence 26, Appl
32	210	16.9	337	15	US-10-282-122A-78161	Sequence 78161, A
33	208	16.8	335	15	US-10-282-122A-55404	Sequence 55404, A
34	207.5	16.7	341	15	US-10-282-122A-67993	Sequence 67993, A
35	207	16.7	441	16	US-10-437-963-111620	Sequence 111620, A
36	206	16.6	337	9	US-09-815-242-10304	Sequence 10304, A
37	206	16.6	337	15	US-10-282-122A-56695	Sequence 56695, A
38	206	16.6	396	15	US-10-425-114-67198	Sequence 67198, A
39	201.5	16.2	421	14	US-10-067-443-4	Sequence 4, Appl
40	201.5	16.2	421	14	US-10-067-443-28	Sequence 28, Appl
41	201.5	16.2	421	15	US-10-649-273-4	Sequence 4, Appl
42	201.5	16.2	421	15	US-10-649-273-28	Sequence 28, Appl
43	201.5	16.2	421	15	US-10-651-722-4	Sequence 4, Appl
44	201.5	16.2	421	15	US-10-651-722-28	Sequence 28, Appl
45	199.5	16.1	350	15	US-10-282-122A-53665	Sequence 53665, A

ALIGNMENTS

RESULT 1
US-10-067-443-22
; Sequence 22, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-22
Query Match 100.0%; Score 1240; DB 14; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e-127; Mismatches 0; Gaps 0;
Matches 239; Conservative 0; Indels 0;

QY 1 LIALVQGVSDFLILGKSLDIAPGMDLKYARLSLIKPEGSGTMSGKAIIEHLAKQGNF 60
DB 29 LIALVQGVSDFLILGKSLDIAPGMDLKYARLSLIKPEGSGTMSGKAIIEHLAKQGNF 88
QY 61 HFDIPELHAKNGCPSTGQVHTDKIINKKKEGIEKQILSSAADIATVQHTMAC 120
DB 89 HFDIPELHAKNGCPSTGQVHTDKIINKKKEGIEKQILSSAADIATVQHTMAC 148
QY 121 HLVKTRHALLFCRQDRLPONNAVIVASGVASNFYIRRALEILTNATQCTLLCPPL 180

Db 149 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNAQTCLLCPPL 208
QY 181 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 239
Db 209 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 267

RESULT 2

US-10-649-273-22
; Sequence 22, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-22

Query Match 100.0%; Score 1240; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 60
Db 29 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 88
QY 61 HFDIKPPLHAHAKCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120
Db 89 HFDIKPPLHAHAKCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 148
QY 121 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNAQTCLLCPPL 180
Db 149 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNAQTCLLCPPL 208
QY 181 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 239
Db 209 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 267

RESULT 3

US-10-651-722-22
; Sequence 22, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 267

; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-22

Query Match 100.0%; Score 1240; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 60
Db 29 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 88
QY 61 HFDIKPPLHAHAKCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120
Db 89 HFDIKPPLHAHAKCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 148
QY 121 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNAQTCLLCPPL 180
Db 149 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNAQTCLLCPPL 208
QY 181 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 239
Db 209 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 267

RESULT 4

US-10-067-443-2
; Sequence 2, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-067-443-2

Query Match 100.0%; Score 1240; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 60
Db 176 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 235
QY 61 HFDIKPPLHAHAKCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120
Db 236 HFDIKPPLHAHAKCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 295
QY 121 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNAQTCLLCPPL 180
Db 296 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNAQTCLLCPPL 355
QY 181 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 239
Db 356 CTDNGIMAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEGASIKVPOLKMEI 414

RESULT 5

US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1

```

; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-649-273-2

```

```

Query Match      100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 LLAIVGVSDFLLLGKSLDIAPGMDLVARSLSLIKHPECSTMGGAIEHLAKQGNRF 60
Db 176 LLAIVGVSDFLLLGKSLDIAPGMDLVARSLSLIKHPECSTMGGAIEHLAKQGNRF 235
Qy 61 HFDIKPPLHAKNCDSPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120
Db 236 HFDIKPPLHAKNCDSPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 295
Qy 121 HLVRTRHAILFCQKRDLLPNNNAVVASGVASNFYIRRALIELTNATQCTLLCPPPL 180
Db 296 HLVRTRHAILFCQKRDLLPNNNAVVASGVASNFYIRRALIELTNATQCTLLCPPPL 355
Qy 181 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
Db 356 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414

```

```

RESULT 6
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-651-722-2

```

```

Query Match      100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 LLAIVGVSDFLLLGKSLDIAPGMDLVARSLSLIKHPECSTMGGAIEHLAKQGNRF 60
Db 176 LLAIVGVSDFLLLGKSLDIAPGMDLVARSLSLIKHPECSTMGGAIEHLAKQGNRF 235

```

```

Qy 61 HFDIKPPLHAKNCDSPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120
Db 236 HFDIKPPLHAKNCDSPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 295
Qy 121 HLVRTRHAILFCQKRDLLPNNNAVVASGVASNFYIRRALIELTNATQCTLLCPPPL 180
Db 296 HLVRTRHAILFCQKRDLLPNNNAVVASGVASNFYIRRALIELTNATQCTLLCPPPL 355
Qy 181 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
Db 356 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414

```

```

RESULT 7
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRF
; ORGANISM: homo sapiens
US-10-067-443-19

```

```

Query Match      97.9%; Score 1213.5; DB 14; Length 439;
Best Local Similarity 90.2%; Pred. No. 3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

Qy 1 LLAIVGVSDFLLLGKSLDIAPGMDLVARSLSLIKHPECSTMGGAIEHLAKQGNRF 60
Db 176 LLAIVGVSDFLLLGKSLDIAPGMDLVARSLSLIKHPECSTMGGAIEHLAKQGNRF 235
Qy 61 HFDIKPPLHAKNCDSPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120
Db 236 HFDIKPPLHAKNCDSPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 295
Qy 99 ---EKQILSSADIAATVQHTMACHLVRTRHAILFCQKRDLLPNNNAVVASGVASN 155
Db 296 CRYEKGQILSSADIAATVQHTMACHLVRTRHAILFCQKRDLLPNNNAVVASGVASN 355
Qy 156 FYIRRALIELTNATQCTLLCPPRLCTDNGIMAMNGIERLRAGILHDIEGIRYEPK 215
Db 356 FYIRRALIELTNATQCTLLCPPRLCTDNGIMAMNGIERLRAGILHDIEGIRYEPK 415
Qy 216 PLGVDISKVEGASIKVPOLKMEI 239
Db 416 PLGVDISKVEGASIKVPOLKMEI 439

```

```

RESULT 8
US-10-649-273-19
; Sequence 19, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05

```

```

; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

```

```

Query Match          97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

QY 1 LLAIVQGVSDFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKGNRF 60
DB 176 LLAIVQGVSDFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI----- 98
DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI FLSKVQINIPGLCLKIAHF 295
QY 99 ---EKQQLSSAADIAATVQHTMACHLVYKTRAILFCQKRDLPONNAVVASGVASN 155
DB 296 CRYEKQQLSSAADIAATVQHTMACHLVYKTRAILFCQKRDLPONNAVVASGVASN 355
QY 156 FYIRRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGILHIDIGIRYEPKC 215
DB 356 FYIRRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGILHIDIGIRYEPKC 415
QY 216 PLGVDISKEVGASIKVPOLKMEI 239
DB 416 PLGVDISKEVGASIKVPOLKMEI 439

```

```

RESULT 9
US-10-651-722-19
; Sequence 19, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

```

```

Query Match          97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

QY 1 LLAIVQGVSDFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKGNRF 60
DB 176 LLAIVQGVSDFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI----- 98
DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI FLSKVQINIPGLCLKIAHF 295

```

```

QY 99 ---EKQQLSSAADIAATVQHTMACHLVYKTRAILFCQKRDLPONNAVVASGVASN 155
DB 296 CRYEKQQLSSAADIAATVQHTMACHLVYKTRAILFCQKRDLPONNAVVASGVASN 355
QY 156 FYIRRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGILHIDIGIRYEPKC 215
DB 356 FYIRRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGILHIDIGIRYEPKC 415
QY 216 PLGVDISKEVGASIKVPOLKMEI 239
DB 416 PLGVDISKEVGASIKVPOLKMEI 439

```

```

RESULT 10
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leibny, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

```

```

Query Match          97.0%; Score 1203; DB 14; Length 414;
Best Local Similarity 97.1%; Pred. No. 3.9e-123;
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 LLAIVQGVSDFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKGNRF 60
DB 176 LLAIVQGVSDFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI EKGQLSSAADIAATVQHTMAC
DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI EKGQLSSAADIAATVQHTMAC 295
QY 121 HLVKTRTRAILFCQKRDLPONNAVVASGVASN FYIRRALEILTNATQCTLLCPPRL 180
DB 296 HLVKTRTRAILFCQKRDLPONNAVVASGVASN FYIRRALEILTNATQCTLLCPPRL 355
QY 181 CTDNGIMIANGIERLRAGILHIDIGIRYEPKPLGVDISKEVGASIKVPOLKMEI 239
DB 356 CTDNGIMIANGIERLRAGILHIDIGIRYEPKPLGVDISKEVGASIKVPOLKMEI 414

```

```

RESULT 11
US-10-094-749-2039
; Sequence 2039, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAL, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO

```

APPLICANT: YAMAMOTO, JUN-ICHI
 APPLICANT: -ISONO, YUUKO
 APPLICANT: HIO, YURI
 APPLICANT: OTSUKA, KAORU
 APPLICANT: NAGAI, KEIICHI
 APPLICANT: IRIE, RYOTARO
 APPLICANT: TAMECHIKA, ICHIRO
 APPLICANT: SEKI, NAOHICO
 APPLICANT: YOSHIKAWA, TSUTOMU
 APPLICANT: OTSUKA, MOTOKUJI
 APPLICANT: NAGAHARI, KENJI
 APPLICANT: MASUHO, YASUHIKO
 TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
 FILE REFERENCE: 084335/0160
 CURRENT APPLICATION NUMBER: US/10/094,749
 PRIOR FILING DATE: 2002-03-12
 PRIOR APPLICATION NUMBER: 60/350,435
 PRIOR FILING DATE: 2002-01-24
 PRIOR APPLICATION NUMBER: JP 2001-328381
 PRIOR FILING DATE: 2001-09-14
 NUMBER OF SEQ ID NOS: 3381
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2039
 LENGTH: 364
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-094-749-2039

Query Match 77.4%; Score 960; DB 15; Length 364;
 Best Local Similarity 98.4%; Pred. No. 1.7e-96;
 Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLGKSLDIAPGMDLVKVARLSLTKHPECSTMGSGKAIIEHLAQGNRF 60
 DB 176 LLAIVGVSDPFLGKSLDIAPGMDLVKVARLPKIKPECSTMGSGKAIIEHLAQGNRF 235
 QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKIMKKEKEGIEKGQILSSADIAATVOHTMAC 120
 DB 236 HFDIKPPLHAKNCDPFTGLQHTVDKIMKKEKEGIEKGQILSSADIAATVOHTMAC 295
 QY 121 HLKRTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPEPL 180
 DB 296 HLKRTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPEPL 355
 QY 181 CTONGIMIA 189
 DB 356 CTONGIMIA 364

RESULT 12
 US-10-424-599-209259
 ; Sequence 209259, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 209259
 ; LENGTH: 445
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_3098C.1.pep
 US-10-424-599-209259

Query Match 27.2%; Score 337; DB 15; Length 445;

Best Local Similarity 38.5%; Pred. No. 6.7e-28;
 Matches 85; Conservative 29; Mismatches 83; Indels 24; Gaps 5;

QY 1 LLAIVGVSDPFLGKSLDIAPGMDLVKVARLSLTKHPECSTMGSGKAIIEHLAQGNRF 60
 DB 194 LLAIVGVSDPFLGKSLDIAPGMDLVKVARLPKIKPECSTMGSGKAIIEHLAQGNRF 248
 QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKIMKKEKEGIEKGQILSSADIAATVOHTMAC 115
 DB 249 SVKFTVPMQKHCNCFSTYAGLTKQVLALESKIDAKIPSSASNGDL-SRADIASFQ 307
 QY 116 HTMAHLVKTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLC 175
 DB 308 RIALVHLERCEALIGWALKEPISIH---LVVSGVASNQYVARLDVVKXGQGLVC 364
 QY 176 PPRRLCTDNGIMIANNGIRLRLAGLIGLIDIEGIRYEPKCP 216
 DB 365 PPRRLCTDNGIMIANNGIRLRLAGLIGLIDIEGIRYEPKCP 395

RESULT 13
 US-10-067-443-3
 ; Sequence 3, Application US/10067443
 ; Publication No. US20030082782A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bristol-Myers Squibb Company
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
 ; FILE REFERENCE: SPINAL CORD, MP-1
 ; CURRENT APPLICATION NUMBER: US/10/067,443
 ; CURRENT FILING DATE: 2002-02-05
 ; PRIOR APPLICATION NUMBER: US 60/266,518
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: US 60/282,814
 ; PRIOR FILING DATE: 2001-04-10
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 463
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-10-067-443-3

Query Match 26.3%; Score 326; DB 14; Length 463;
 Best Local Similarity 33.6%; Pred. No. 1.1e-26;
 Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 1 LLAIVGVSDPFLGKSLDIAPGMDLVKVARLSLTKHPECSTMGSGKAIIEHLAQGNRF 60
 DB 221 LLAIVGVSDPFLGKSLDIAPGMDLVKVARLPKIKPECSTMGSGKAIIEHLAQGNRF 275
 QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKIMKKEKEGIEKGQILSSADIAATVOHTMAC 120
 DB 276 SVKFTVPMQKHCNCFSTYAGLTKQVLALESKIDAKIPSSASNGDL-SRADIASFQ 325
 QY 121 HLKRTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPEPL 180
 DB 326 HLBKRCERAILDIALR---LEPSIKHNVISGVASNFYIRLNNIVENKRLKLVCPPEPL 382
 QY 181 CTONGIMIANNGIRLRLAGLIGLIDIEGIRYEPKCP 224
 DB 383 CTONGIMIANNGIRLRLAGLIGLIDIEGIRYEPKCP 432
 QY 225 VGEA 228
 DB 433 RSEA 436

RESULT 14
 US-10-649-273-3
 ; Sequence 3, Application US/10649273
 ; Publication No. US20040043407A1
 ; GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:10:32 ; Search time 3803.22 Seconds
(without alignments)
3045.000 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LILALVGVSPFLILGKSLDI.....DISKEVGESIKVPQLKMERI 239

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 segs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+g2n.model -DEV=x1h
-Q=/cgn21/USPTO.spool/US10649273/runat.14022005.114702.16399/app.query.fasta_1.1429
-DB=GenEmbl -OFMT=faastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOEXT=45
-UNITS=bites -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10649273 @CGN_1.1 8655 @runat.14022005.114702.16399 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NRG SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hrg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length DB	ID	Description
1	1240	100.0	1416	AR541929 Sequence
2	1240	100.0	1526	AR428809 Sequence
3	1240	100.0	1908	BC011904 Homo sapi
4	1240	100.0	2197	AR428803 Sequence

5	1213.5	97.9	1387	6	AR428808	AR428808 Sequence
6	1213.5	97.9	1387	6	HS4295148	AR295148 Homo sapi
7	1203	97.0	1245	6	AX664697	AX664697 Sequence
8	1203	97.0	1820	6	AX664695	AX664695 Sequence
9	1086	87.6	1844	10	BC058172	BC058172 Mus muscu
10	1079	87.0	1017	10	BC038910	BC038910 Mus muscu
11	1059	85.4	2208	6	AX713716	AX713716 Sequence
12	1059	85.4	2208	6	AK055441	AK055441 Homo sapi
13	983	79.3	1546	10	BC078974	BC078974 Rattus no
14	944.5	76.2	1109	10	BC051211	BC051211 Mus muscu
15	809	65.2	860	5	BX330694	BX330694 Gallus ga
16	806	65.0	1522	5	BX934991	BX934991 Gallus ga
17	803	64.8	1558	5	BX930963	BX930963 Gallus ga
18	725	58.5	14364	6	AR428807	AR428807 Sequence
19	725	58.5	84115	9	AC013468	AC013468 Homo sapi
20	666	53.7	249601	2	AC114153	AC114153 Rattus no
21	666	53.7	308652	2	AC121478	AC121478 Rattus no
22	646.5	52.1	256751	10	AC122925	AC122925 Mus muscu
23	610	49.2	875	6	CQ721898	CQ721898 Sequence
24	384.5	31.0	1474	3	AK113378	AK113378 Clona int
25	369.5	29.8	121251	5	AL591593	AL591593 Zebrafish
26	356	28.7	20021	10	AF367967S1	AF367967 Mus muscu
27	356	28.7	179252	10	AF131205	AF131205 Mus muscu
28	345	27.8	117322	5	AL672217	AL672217 Zebrafish
29	332.5	26.8	1576	3	AY051882	AY051882 Drosophila
30	332.5	26.8	1601	6	CQ606432	CQ606432 Sequence
31	332.5	26.8	3656	6	CQ606431	CQ606431 Sequence
32	332.5	26.8	14679	2	AC018262	AC018262 Drosophila
33	332.5	26.8	180263	3	AC010671	AC010671 Drosophila
34	332.5	26.8	207432	3	AE003513	AE003513 Drosophila
35	324	26.1	1443	8	AY024338	AY024338 Arabidops
36	324	26.1	1474	8	AY117283	AY117283 Arabidops
37	324	26.1	1567	8	AY063864	AY063864 Arabidops
38	322	26.0	1557	8	AY084577	AY084577 Arabidops
39	311.5	25.1	1672	8	AK070912	AK070912 Oryza sat
40	305	24.6	110000	2	AP006501	AP006501 Oryza sat
41	262.5	21.2	333800	1	SME591792	SME591792 Stenohizo
42	260	21.0	1718	8	AK099965	AK099965 Oryza sat
43	257	20.7	349354	1	BX640416	BX640416 Bordetella
44	254	20.5	260271	1	AE017258	AE017258 Wolbachia
45	253	20.4	349116	1	AP003003	AP003003 Mesorhizo

ALIGNMENTS

RESULT 1	AR541929	1416 bp	DNA	linear	PAT 08-OCT-2004
LOCUS	AR541929	Sequence 177 from patent US 6743619.			
DEFINITION	AR541929				
ACCESSION	AR541929				
VERSION	AR541929.1	GI:53934009			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1416)				
AUTHORS	Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Asundi,V., Ren,F., Zhang,J., Zhao,Q.A., Yang,Y., Xue,A.D., Wehrman,T., Wang,J.-R., Wang,D. and Drmanac,R.T.				
TITLE	Nucleic acids and polypeptides				
JOURNAL	Patent: US 6743619-A 177 01-JUN-2004;				
FEATURES	Location/Qualifiers				
source	1..1416				
ORIGIN	/organism="unknown"				
	/mol_type="genomic DNA"				

Alignment Scores:
Pred. No.: 2,242-113
Score: 1240.00
Percent: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Length: 1416
Matches: 239
Conservative: 0
Mismatch: 0
Indels: 0

DB: 6 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AR541929 (1-1416)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
 DB 586 CTGTTGGCATTAGTTCAGAGAGTTTCAGATTTTCGCTTCTTGAAAGCTTTGGACATA 645

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAspLeuSerLeuIleLysHisProGlu 40
 DB 646 GCACACAGGTGACATGCTTGACAGGTGGCAAGAACCTTTCTTAATAAATCAATCCAGAG 705

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
 DB 706 TGCCTCCACCATGAGTGGTGGGAAGCCATAGAACCTTTGGCCAAACAGGAATAGATT 765

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
 DB 766 CATTTCGACATCAAACTCCCTTGCAATGCTTAATAAATTTGATTTTCTTTTACTGGA 825

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGluGlyIleGluLys 100
 DB 826 CTTCAACACGTTACGTATGATAAATAATGAAAAAGAAAAAGAGAGTATGAGAG 885

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
 DB 886 GGGCAATTCCTGCTTCGACGACGACATTCCTGCGCACGACGACACATGCGCATGT 945

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
 DB 946 CATCTTGGAAAAAGAACACATCGGCTATTCTGTTTGTGAGCAGAGACTGTACT 1005

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheTyrlleArgArg 160
 DB 1006 CAAATATATGAGATCTGGTGCATCTGGTGTGTCGAAAGTACCTCTATATCCGACGA 1065

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
 DB 1066 GCTCTGGAATTTTAAACAACGCAACACAGTGCATTTGTTGTGTCTCTCCACACTA 1125

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
 DB 1126 TGCACGTATATGCAATTATGATTCATGATGCAATGGAATGGAAGCTACGTGCTTG 1185

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrluProLysCysProLeuGlyValAsp 220
 DB 1186 GGCATTTTAACTAGCATAGAGGCAATCCGCTATGAAACCAAAATGTCTCTTGAGTAGAC 1245

QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
 DB 1246 ATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAAATGGAGATA 1302

RESULT 2
 AR428809 1526 bp DNA linear PAT 18-DEC-2003
 LOCUS Sequence 23 from patent US 6642041.
 DEFINITION AR428809
 ACCESSION AR428809
 VERSION AR428809.1 GI:40188595
 KEYWORDS

SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1526)
 AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclof, F.
 TITLE Polynucleotides encoding a novel metalloprotease, MP-1
 JOURNAL Patent: US 6642041-A 23 04-NOV-2003;
 FEATURES location/Qualifiers
 1..1526
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Alignment Scores:

Pred. No.: 2,456-113 Length: 1526
 Score: 1240.00 Matches: 239
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AR428809 (1-1526)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
 DB 85 CTGTTGGCATTAGTTCAGAGAGTTTCAGATTTTCGCTTCTTGAAAGCTTTGGACATA 144

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAspLeuSerLeuIleLysHisProGlu 40
 DB 145 GCACACAGGTGACATGCTTGACAGGTGGCAAGAACCTTTCTTAATAAATCAATCCAGAG 204

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
 DB 205 TGCCTCCACCATGAGTGGTGGGAAGCCATAGAACCTTTGGCCAAACAGGAATAGATT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
 DB 265 CATTTCGACATCAAACTCCCTTGCAATGCTTAATAAATTTGATTTTCTTTTACTGGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGluGlyIleGluLys 100
 DB 325 CTTCAACACGTTACGTATGATAAATAATGAAAAAGAAAAAGAGAGTATGAGAG 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
 DB 385 GGGCAATTCCTGCTTCGACGACGACATTCCTGCGCACGACGACACATGCGCATGT 444

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
 DB 445 CATCTTGGAAAAAGAACACATCGGCTATTCTGTTTGTGAAACAGAGACTGTACT 504

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheTyrlleArgArg 160
 DB 505 CAAATATATGAGATCTGGTGCATCTGGTGTGTCGAAAGTACCTCTATATCCGACGA 564

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
 DB 565 GCTCTGGAATTTTAAACAACGCAACAGTGCATTTGTTGTGTCTCTCCACACTA 624

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
 DB 625 TGCACGTATATGCAATTATGATTCATGATGCAATGGAATGGAAGCTACGTGCTTG 684

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrluProLysCysProLeuGlyValAsp 220
 DB 685 GGCATTTTAACTAGCATAGAGGCAATCCGCTATGAAACCAAAATGTCTCTTGAGTAGAC 744

QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
 DB 745 ATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAAATGGAGATA 801

RESULT 3
 BC011904 1908 bp mRNA linear PRI 23-DEC-2003
 LOCUS Homo sapiens O-6-ialoglycoprotein endopeptidase-like 1, mRNA (cDNA
 DEFINITION clone MGC:20293 IMAGE:4121450), complete cds.
 ACCESSION BC011904
 VERSION BC011904.2 GI:40225818
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1908)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Martusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stadelton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McErihan, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Senger, A., Schein, J.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932
2 (bases 1 to 1908)
Strausberg, R.
Direct Submission
Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
On Dec 19, 2003 this sequence version replaced gi:15080281.
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Ahter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Taurgeon, C., Vogt, J.B., Walker, M.A., Weethey, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 28 Row: 1 Column: 22.
Location/Qualifiers
1. 1908
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:20293 IMAGE:4121450"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_id="NIH MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. 1908
/gene="OSGEP1"
/db_xref="LocusID:64172"
104. 1348
/gene="OSGEP1"
/codon_start=1
/product="OSGEP1 protein"
/protein_id="AAH11904.1"
/db_xref="GI:15080282"
/translation="MLITKTAGVFKKRVKVEFLRSPNRPGLTFLAKIVLGIET
SCDDTAAYVDEGNVIGEAISQTEVLEKTCGIVPAAGQAPENIQRIVQALASAS
GVSPSDLSAIAITTKPGIALSLVGGLSFLDVLGQKKPFIPIHMEALITRLINK

VERPVLVLLISGCHCLLAVGVSDFLLEKSLDIARGMIDRVARRLSIKPEGST
MSGKAIIEHAKQNFEPDIPKTHANKQCSFPGQIVTDKIKMKEKEGIEKG
GLSSAADIATATQFMACHLVKTRTALIFCGRDLIPONNAVLSGGAANFYR
RLREITLNTQCTLCPPRLCTPDNIMAMNIEHRLRAGILHIEGRVPEKPL
GVDISVEVEASIKVPQLMET"
215. 1111
/gene="OSGEP1"
/note="Peptidase M22; Region: Glycoprotease family"
/db_xref="CD:pfam00814"

misc_feature
Alignment Scores:
Pred. No.: 3,19e-113 Length: 1908
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BC011904 (1-1908)

QY 1 LeuLeuAlaLeuValGlnGlyValSerApphLeuLeuGlnGlySerLeuAaple 20
DB 629 CTGTTGGCATTGATTGACAGAGATTCTCTCTTGGAAAGTCTTGGACATA 688
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLysSerLeuLeuLysHisProGlu 40
DB 689 GACCCAGGACATGCTTGACAGAGGTGGCAAGACCTTCTTAATAAATCCAGAG 748
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 749 TCGTCCACATGATGCTGGGAAAGCCATAGACATTGGCCAAACAGAAATAGTTT 808
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheThrGly 80
DB 809 CATTTTGACATCAAACTCCCTTCATCATGCTAAAAATGTTCTTTTCTTACATGGA 868
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyLys 100
DB 869 CTTCAACACGTTATGATATAAATAAATAAATAAAGAAAGAAAGGATATTGAGAG 928
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
DB 929 GGGCAAAATCTGCTTCAAGCAGACATTTGCTGCCACATACACACAAATGGCATGT 968
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 989 CATCTTGAAAAGAACACATCGGGCTATCTGTTGTAAAGCAGAGACTGTAACT 1048
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArg 160
DB 1049 CAATAAATGACAGTATGCTGTGATCGTGGTGGCAAGTAATCTTATATCCACAGA 1108
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1109 GCTCTGGAATTTTAAACAACGACACAGTGCATTTGTTGTCTCTCTCCACACTA 1168
QY 181 CysThrAspAsnGlyIleMetIleAlaArgAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
DB 1169 TCACATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 1228
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrlleArgProLysCysProLeuGlyValAsp 220
DB 1229 GGCATTTTACATGACATTAAGAGCATCCCTATGAACCAAAATGTCCTCTTGAAGTAGAC 1288
QY 221 HisSerLysGlnValGlnGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239
DB 1289 AATATCAAAAGAGTGGAGAGCTTCATTAAGTACCACAATTAATAAATGAGATA 1345
RESULT 4
AR428803
LOCUS AR428803 2197 bp DNA linear PART 18-DEC-2003
DEFINITION Sequence 1 from patent US 6642041.

ACCESSION AR428803
VERSION AR428803.1 GI:40188589
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 2197)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;
FEATURES
Location/Qualifiers
1..2197
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 3.77e-113 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AR428803 (1-2197)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 756 CTGTTGGCATTAGTTCAGAGAGATTTCAGATTTCGCTTCCTGGAAAGCTTTGGACATA 815

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 816 GCACCAAGTGACATGCTTGCACAGAGTGCGCAAGAAACCTTTCTTAATAAATCCACAG 875

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuValLysGlnGlyAsnArgPhe 60
DB 876 TGCTCCACCATGAGTGTGGGAAAGCCATAGAACTTTGGCCAAACAGAAATAGATT 935

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 936 CATTGTGACATCAAACTCCCTGCATCATGCTAAAAATGGAATTTTCTTTACTGGA 995

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyLeu 100
DB 996 CTTCAACACGTTACGATTAATAATATGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 1055

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1056 GGGCAAAATCCTGTCTTCAGCAGCAGCAATGCTGCACAGTACAGCAGCAGCATGGCATGT 1115

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1116 CATCTTGGAAGAAAGACATCGGGCTATCTGTTTGTAAACAGAGAGACTGTAACTCT 1175

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaLysAsnPheThrIleArgArg 160
DB 1176 CAATAATATGAGATGCTGTCATCTGTCGTGTCGCAAGTAACCTTCAATATCCGCA 1235

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1236 GCTCTGGAAATTTTAAACAAAGCAGCAGCAGCTTTGTTGTCCTCCCTCCAGACAT 1295

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1296 TGCACTGTAATGCGATTATGATTCATGAGATGCTATGAAAGACTAGCTCGCTGG 1355

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgGluProLysCysProLeuGlyValAsp 220
DB 1356 GGCATTTTAACTAGACATGAAAGCATCCGCTATGAAACCAAAATGTCCTTGAAGTAGAC 1415

QY 221 IleSerLysGluValGlyGluLysSerIleLysValProGlnLeuLysMetGluIle 239
DB 1416 ATATCAAAAGAGAGTTGGAGAGCTTCCATTAAGTACCAATTTAAATGAGAGATA 1472

RESULT 5
LOCUS AR428808 1387 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 21 from patent US 6642041.
ACCESSION AR428808
VERSION AR428808.1 GI:40188594
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1387)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;
FEATURES
Location/Qualifiers
1..1387
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 9.37e-111 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 97.86% Mismatches: 1
Query Match: 97.86% Indels: 25
Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x AR428808 (1-1387)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 549 CTGTTGGCATTAGTTCAGAGAGATTTCAGATTTCGCTTCCTGGAAAGCTTTGGACATA 608

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 609 GCACCAAGTGACATGCTTGCACAGAGTGCGCAAGAAACCTTTCTTAATAAATCCACAG 668

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuValLysGlnGlyAsnArgPhe 60
DB 669 TGCTCCACCATGAGTGTGGGAAAGCCATAGAGCAATTTGGCCAAACAGAAATAGATT 728

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 729 CATTGTGACATCAAACTCCCTGCATCATGCTAAAAATGGAATTTTCTTTACTGGA 788

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyLeu 98
DB 789 CTTCAACACGTTACGATTAATAATATGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 848

QY 98 ----- 98

DB 849 ATTAGTAAGTGAACAGATTAATATTCTGATGTCGCTAAATAATAGCTGCTATTTC 908

QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
DB 909 TCGAGGTATGAGAAAGGGGCAAAATCCGCTTCAGCAGCAGCACTTCGCAAGTAACAG 968

QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB 969 CACACATGCGATGTCATCTTGAAAGAAACATCGGGCTATCTGTTTGTAAAGCAG 1028

QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaLysAsn 155
DB 1029 AGAGACTTGTTCCTCAAAATATGCAAGTACGTTGCAATCTGTGTGTGCGCAAGTAAC 1088

QY 156 PheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175
DB 1089 TTCTATATCCGAGACACTCTGGAAATTTTAAACAAAGCAGCAACAGTGCACTTTGTTGTGT 1148

QY 176 ProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArg 195

Db 1149 CCTCCTCCAGACTATGCACTGATTAATGCGATTGATGCAATGATGATTTGGAAGA 1208
QY 196 leuargalaglyleuglylleuuhisaapilleglyllyeargtyrctuprolyscys 215
Db 1209 CTAGCTGGTGGCTTGGGCAATTTTACATGACATGAGAGCATCCGCTATACCAAAATGT 1268
QY 216 ProleuglyvalaspiIeserlysglyvalaglyualaseriallelysvaIproglneu 235
Db 1269 CCTCTGGAGTAGACATATCAAAAGAGTTGGAGAGGCTTCATTAAGTACACACATTA 1328
QY 236 LysMetGluIle 239
Db 1329 AAAATGAGATA 1340
RESULT 6
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000
LOCUS HSA295148 Homo sapiens mRNA for putative sialoglycoprotease type 2.
ACCESSION AJ295148.1 GI:11071726
VERSION AJ295148.1 GI:11071726
KEYWORDS metalloproteinase; sialoglycoprotease.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 Chen J.M., Fortunato M. and Barrett A.J.
TITLE Cloning and sequencing of a second human putative
REFERENCE 2 (bases 1 to 1387)
AUTHORS Chen J.M.
JOURNAL Direct Submission
TITLE Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology
JOURNAL Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,
JOURNAL UNITED KINGDOM
FEATURES
Source 1. 1387 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="uterus"
24. 1343
/function="metalloproteinase of family M22"
/codon_start=1
/product="putative sialoglycoprotease type 2"
/protein_id="CAC14666.1"
/db_xref="GI:11071727"
/db_xref="GOA:Q9H4B0"
/db_xref="UniProt/TREMBL:Q9H4B0"
/translation="MLILTKTAGVFPFKSKRVYELRSPNHPBTLFLAKIVLGIEI
SCDDTAAYVDETNVLEBAIHSQTEVHLKTGIVPPAAQQLHRENIORIVQEMLSAS
GVSPDLSAIAITTKPGIALSLIGVLSFDLVGQLKCPPIPIHMEALATIRLTK
VEPFVLVLTISGHCIALVQGVSDFLIGSLDIPGMDLKVARRSLIKHPECS
MSGKALEHAIAKGNRPFPDIKPKLHAKNGDSEPTGQHYTDKIIMKKKEGFI
SKRQINIRPCLCTIAAHFRCYKEGQIISAAADIATYQHTMACILVKTRAILFCK
QRDLIPQNNVNLVLSGVASNFYIRRAIELITNAITCTLLCPPELCTDNGIMTAWG
IERLRGIGILHDIEGIRYBKCPGVDISKEVGEASIKVPLKMEI"
ORIGIN
Alignment Scores:
Pred. No.: 9.37e-111 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: 9 Gaps: 1
US-10-649-273-2_COPY_176_414 (1-239) x HSA295148 (1-1387)
QY 1 leuenuIleuValGInGlyValSerAspHeuLeuGlyLysSerIleuaspIle 20
Db 549 CTGTGGCATTTAGTACAGAGGATTTTCGCTTCTTGAGAAAGCTTTGGACATA 608

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerIleuIleYshIsprogly 40
Db 609 GCACCGAGTGACATGCTTTGACAAAGTGGCAAGAGACTTTCTTAATAAATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisleuAlaLysGInGlyAsnArgPhe 60
Db 669 TGCTCCACCATGAGTGGTGGAAAGCCATAGACATTTGGCCAAACAGGAAATAGATT 728
QY 61 HisPheAspIleIleYsProProleuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
Db 729 CATTTTGACATCAAACTCCCTGCATCATGCTTAAAAATTTGATTTTCTTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlyGluGlyIle----- 98
Db 789 CTTCAACAGCTTACTGATTAATAATTAATGAAGAAAGGAAAGAGAGATTAATTTCTA 848
QY 98 ----- 98
Db 849 ATTAGTAAGTTGAACAGATAATATATTCCTGATTTGGCTTAATAAATAGCTGCTATTTC 908
QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
Db 909 TCGAGGTATGAGAGGGGCAAAATCCTGCTTCAGCAGACATTCCTGCACAGTACAG 968
QY 116 HisThrMetAlaCysHisleuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
Db 969 CACACATGACATGTCATCTTGTAAGAAAGACATCGGGCTATTTCTGTTTAAAGCAG 1028
QY 136 ArgAspLeuLeuProGlnAsnAlaValleuValAlaSerGlyGlyValAlaSerAsn 155
Db 1029 AGAGACTTATACCTCAAAATTAATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1088
QY 156 PheTyrlIleArgArgAlaLeuGlyIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCys 175
Db 1089 TTCATATCCGACGAGCTCTGAAATTTTAAACAAAGCAACACAGACACTTTGTTGTGT 1148
QY 176 ProProProArgLeuLeuThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArg 195
Db 1149 CCTCCTCCAGACTATGCACTGATTAATGCGATTGATGCAATGATGATTTGGAAGA 1208
QY 196 leuargalaglyleuglylleuuhisaapilleglyllyeargtyrctuprolyscys 215
Db 1209 CTAGCTGGTGGCTTGGGCAATTTTACATGACATGAGAGCATCCGCTATACCAAAATGT 1268
QY 216 ProleuglyvalaspiIeserlysglyvalaglyualaseriallelysvaIproglneu 235
Db 1269 CCTCTGGAGTAGACATATCAAAAGAGTTGGAGAGGCTTCATTAAGTACACACATTA 1328
QY 236 LysMetGluIle 239
Db 1329 AAAATGAGATA 1340
RESULT 7
AX664697 1245 bp DNA linear PAT 22-MAR-2003
LOCUS AX664697 Sequence 6 from Patent WO02074960.
ACCESSION AX664697
VERSION AX664697.1 GI:29164457
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Leidy, K.R., Kapeller-Libermann, R. and Gluckmann, M.
TITLE 38650, 28472, 5495, 65507, 81368 and 14354 methods and compositions
JOURNAL of human proteins and uses thereof
PARENT: WO 02074960-A 5 25-SEP 2002;
MILLENNIUM Pharmaceuticals, Inc. (US)
FEATURES
Source 1. 1245 Location/Qualifiers
/organism="Homo sapiens"

ORIGIN /mol_type="unassigned DNA"
/db_xref="taxon:9606"

Alignment Scores:

Pred. No.: 9.1e-110 Length: 1245
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AX664697 (1-1245)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspLe 20
DB 526 CTGTTGGATTAGTTCAGAGGAGTTTCAGATTTTCCTTCTTGAAAGCTTTTGACATA 585
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 586 GCACCAAGTGCATGCTTGACAGAGTGCAGACACATTTCTTAAATAAACATCCAGAG 645
QY 41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaLysGlnGlyValAsnArgPhe 60
DB 646 TGCTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATATGATTT 705
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 706 CATTTTGACATCAAACTCCCTTCATCATGCTAAAAATGGAATTTTCTTTTACTGGA 765
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100
DB 766 CTTCAACGCTTACGATTAATAATATGAAAACAGAAAACAGAGAGAGCTTTGAGAG 825
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleArgValGlnHisThrMetAlaCys 120
DB 826 GGGCAAAATCCTGTCTTCAGCAGACACATTTGCTGCACAGTACAGACACAAATGGCAT 885
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 886 CATCTTGGAAGAAAGACATCGGGCTATTCGTTTGTGACAGAGAGACTGTTTACCT 945
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
DB 946 CAAATATATGACATGATCGATGATCTGATGATGATGATGATGATGATGATGATGAT 1005
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1006 GCTCTGGAAATTTTAAACAAAGCACAAGTGCATTTGTTCTGCTCTCCAGACATA 1065
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgPheGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1066 TGCACATGATATGCGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1125
QY 201 GlyIleLeuHisAspIleGlyIleArgGlyIleArgGlyIleArgGlyIleArgGlyIle 220
DB 1126 GGCATTTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1185
QY 221 IleSerLysGluValGlyValAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1186 ATATCAAAAGAGAGTTGAGAGAGCTTCCATTAAGATACCAATTTAAATGAGAGATA 1242
```

RESULT 8
AX664695 1820 bp DNA linear PAT 22-MAR-2003
LOCUS Sequence 4 from Patent WO02074960.
DEFINITION AX664695
ACCESSION AX664695.1 GI:29164455
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
of human proteins and uses thereof
JOURNAL Patent: WO 02074960-A 4 26-SEP-2002;
Millennium Pharmaceuticals, Inc. (US)
FEATURES Location/Qualifiers

source

1. 1820
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
146..1390
/note="unnamed protein product"

CDS

/codon_start=1
/protein_id="CAD80044.1"
/db_xref="GI:29164456"
/translation="MLILTKTAVGFPPKSKRYVEPLRSFNHPGTLPLKIVLGIET
SCDDTAAAVVDEKTVLGEALHSQTEVHLKGTGIVPPAQQOHLRENIQIVQALAS
GVSPELISAITTIKRGALSLGVLSFSLQVGQKKEFPIPHHEAAHLITRLTK
VRFPLVILISGHCILALVQGVSPFLKSLIDLRGMDLVARSLIKHPEST
MSGKRIETLAKQGRFPHDIKPLHNAKCDPSFGLQHTVDKXNENKQSEIEKG
QILSSADIDATVQHTMACHLYKTRHAILFKQKRDLPQNNAVLYVSGVSNPFR
RALRLITNATVQCTLCPPRLCTDNGIMTAMGIEKRLRAGLGLHDI EGIRPEKCP
GVDISKVEGASIKVPLKMEI"

ORIGIN

Alignment Scores:

Pred. No.: 1.43e-109 Length: 1820
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AX664695 (1-1820)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspLe 20
DB 671 CTGTTGGATTAGTTCAGAGGAGTTTCAGATTTTCCTTCTTGAAAGCTTTTGACATA 730
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 731 GCACCAAGTGCATGCTTGACAGAGTGCAGACACATTTCTTAAATAAACATCCAGAG 790
QY 41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaLysGlnGlyValAsnArgPhe 60
DB 791 TCCTCCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 850
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 851 CATTTTGACATCAAACTCCCTTCATCATGCTAAAAATGGAATTTTCTTTTACTGGA 910
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100
DB 911 CTTCAACGCTTACGATTAATAATATGAAAACAGAAAACAGAGAGAGTATGAGAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleArgValGlnHisThrMetAlaCys 120
DB 971 GGCATTTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1030
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1031 CATCTTGGAAGAAAGACATCGGGCTATTCGTTTGTGACAGAGAGCTGTTTACT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
DB 1091 CAAATATATGACATGATCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1150
QY 161 AlaLeuGlnIleLeuThrAsnAlaIleArgGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1151 GCTCTGGAATTTTAAACAAAGCACAAGTGCATTTGTTGTGCTCTCCAGACATA 1210
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgPheGlyIleGluArgLeuArgAlaGlyLeu 200
```


Db		1107	CACCTTGGAAAAGAAACAATCGTCGCTACTTTCTGTTTTCCAGCAGAAAAATTGTCACT	1166	
Oy		141	GlnAsnAnaIaValLeuValAlaSerGlyGlyValAlaSerAsnPhenTyrlleArgArg	160	
Db		1167	CCACCTAACGCAGATATATAGTATGTAATCTTGAGAGGTGTTCGAAGTAACCTTATCATCCGAAAA	1226	
Oy		161	AlaIeuGluIleLeuThrAsnAlaThrGlnCystrIleLeuCyProProColArgLeu	180	
Db		1227	GCATTGGAAATATGTCGCAAAATGCCAAGCAGACGAGTTCGTGTGCCCCCTCCAGAAGCTG	1286	
Oy		181	CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu	200	
Db		1287	TGCATTCAGCATGGCATATATATTCATTCAGAAATGCAATTGAAGATTACGTCGCTGCTG	1346	
Oy		201	GlyIleLeuHisAspIleGluGlyIleArgTyrglyProIyCySProIeuGlyValAsp	220	
Db		1347	GGCGTTTATCATGATGTGTGAAGACATCCGATATGAAACCAAATATGCTCCTTGAAATAAC	1406	
Oy		221	IleSerIySGluValGlyGlyAlaSerIleIySValProGlnLeuIyMetGluIle	239	
Db		1407	ATATCCAGAGAAAGTTCAGAAAGCTGCCATTAAGTACCAGCATTTAAATATGCACTT	1463	
RESULT 10					
LOCUS	BC038910	1017 bp	mRNA	linear	
DEFINITION	Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA				
ACCESSION	clone IMAGE:5053559), partial cde.				
VERSION	BC038910				
KEYWORDS	BC038910.1 GI:24433548				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1017)				
	Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshuler, S.F., Zeeberg, B., Bluetow, K.H., Schaefer, C.F., Bat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapellato, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schepert, T.E., Brownstein, M.J., Ueding, T.B., Tothylki, S., Caranci, P., Prange, C., Rana, S.S., Loquillano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., Mobwan, P.J., McAlexander, K.U., Malek, J.A., Gunaratne, P.H., Richards, S., Wiley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Holly, S.W., Villation, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Heitton, B., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, R., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalka, U., Smalins, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 22388257 12477932 2 (bases 1 to 1017) Straussberg, R. Direct Submission Submitted (25-OCT-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: gcgabs-remail.nih.gov Tissue Procurement: Jeffrey B. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL) DNA Sequencing by: Genome Science Centre,				
REMARK					
COMMENT					

```

BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca
Seven Jones, Jennifer Aaano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
San Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Strot,
Michael Thorne, Miranada Tsai, Nataaja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Maria.

Clone distribution: MCC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.lnln.gov
Series: IRAC Plate: 86 Row: f Column: 12.

Location/Qualifiers
1. 1017
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053559"
/tissue_type="Liver, normal, 5 month old male mouse."
/clone_id="NCI CGAP_L19"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ALIGNMENT SCORES:
Pred. No.: 1.48e-97 Length: 1017
Score: 1079.00 Matches: 205
Percent Similarity: 91.63% Conservative: 14
Best local Similarity: 85.77% Mismatches: 20
Query Match: 87.02% Indels: 0
DB: 10 Gaps: 0

```

```

US-10-649-273-2_COPY_176_414 (1-239) x BC038910 (1-1017)
QY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIle 20
DB      110 CTGTGGCATTTAGTCCAGAGGTGTTCCGATTTCCCTGCTCTGGGAAGTCTTTGGACATA 169
QY      21 AlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuLeuIleYsHisProGlu 40
DB      170 GCACACAGGCCACATGCTTGGACAAAGTGGCAAGAGACATTTCTTTAAACAACATCCAGAA 229
QY      41 CysSerThrMetSerSerGlyIysValAlaIleGlnHisLeuAlaIysGlnIysAsnArgPhe 60
DB      230 TGTTCACAAATGAGTGTGGAAAAAGCTATAGAACAGTGGCCCAAGACGGAAATAGATTC 289
QY      61 HisPheAspIleYsProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGly 80
DB      290 CATTTCATCTTCATCAATCCATCACTATGACAGAAATGCTTAGAATGGCATTTTCTTCAACGGGA 349
QY      81 LeuGlnHisValThrAspIysIleIleMetIysIysGluIysGluGluGlyIleGluIys 100
DB      350 CTTCAACATTTACTGATGTAAGCTAATACACACAGAAAAAGAGAGGATTTGAGAG 409
QY      101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
DB      410 GGGCAATTCCTGCATCATCAGCTGCAGACATTTGCTCGCTCGGTACGACATGCACACGCTGC 469
QY      121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnIleGluAspLeuPro 140
DB      470 CACCTTCGAAAAAGAACACATCCGCGTATTCGTTTGGCAAGCAAGAAAAATTGGCTCTCT 529
QY      141 GluAsnAsnAlaValLeuValAlaSerGlyIysValAlaSerAsnPheArgIleArgArg 160
DB      530 CCACGTACACAGATTAAGTTGATCTGGAGAGGTGGCAAGTAACTGTACATCCAGAAA 589
QY      161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB      590 GCATTGAAATTTGTCGAAATGCAACGCAAGTGCACGTTGTGTGTCCACTCCACAGACTG 649

```

QY 181 CysThraspaanglyllewetiellaatrpanglylleuargleuarglaaglyleu 200
 Db 650 TGCACTGACATGCGATCATGATTCATGAAATGGAATTAAGTCTGCTGCTTG 709

QY 201 G1yleleuHieapllieguglylleargtyrlyuprolvscyeproleuglyvalasp 220
 Db 710 GGGCTTTACATGATGAGAAAGACATCCGATATGAAACAAATGTCCTCTTGAGAGAC 769

QY 221 lleserlysgluvalglylualaserllelyvalproglneuleyswetgulle 239
 Db 770 ATATCCAGAGAGATTGCAAGACCTCCATTAAGTACCCGATTAAATGCGACTT 826

RESULT 11
 AX713716 2208 bp DNA linear PAT 15-APR-2003
 LOCUS Sequence 400 from Patent EP1293569.
 DEFINITION AX713716
 VERSION AX713716.1 GI:29886642
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Chordata; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
 1 Iosagi,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Negahari,K. and Masuhio,Y.
 Full-length cDNAs
 Patent: EP 1293569-A 400 19-MAR-2003;
 Helix Research Institute (JP) ; Research Association for Biotechnology (JP)
 FEATURES
 source location/Qualifiers
 1..2208
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores:
 Pred. No.: 3.57e-95 Length: 2208
 Score: 1059.00 Matches: 211
 Percent Similarity: 88.28% Conservative: 0
 Best Local Similarity: 88.28% Mismatches: 4
 Query Match: 85.40% Indels: 24
 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x AX713716 (1-2208)

QY 1 leuenualeuvalgnglyvalseraphelenuleuglylyeserleuaspille 20
 Db 869 CTGTGTGCATTAAGTTCAGAGATTTCGCTTCTTGGAAGCTTTTGACATTA 928

QY 21 Alaproslyaspmetleuaspylevalalargargleuserleulleysihisproglu 40
 Db 929 GCACCAAGTACATCTTGACAGAGTGCAGAGACATTCCTTTAATTAACATCAAG 988

QY 41 CysserThretserglylylyalalalegluhsleualalyegnglylyasnarphe 60
 Db 989 TGCTCCACATAGATGAGTGGGAAAGCCATGACATTTGGCCAAACAGAAATGATTT 1048

QY 61 HiephaspillelyproproleuHiehsialalyasancysapppheserphethrgly 80
 Db 1049 CATTTCACATCAACCTCCCTTCATCATCTAATAAATTCGATTTCTTTTACTGCA 1108

QY 81 leuGluHieValThraspylelleleuetylysglyugnglyllegluys 100
 Db 1109 CTTCAACACGTTACTGATTAATAATTAATGAAAAGGAAAAGGAAAGATTTGAGAG 1168

QY 101 G1yGlnileuuseriseralealaspillealathrvalglnhietHmetAlaCys 120
 Db 1169 GGGCAATTCCTGCTTTCAGCAGCAGCATTCCTCCACATGACACATGACATCT 1228

QY 121 HieleuVallysaargthrhisaargalalleuPheCyslysglnargaspLeuPro 140
 Db 1229 CATTCTGTGAAAAGAACATCGGGCATTCCTGTTTGTAAAGCAGACAGACTTGTACT 1288

QY 141 GlnasnasnalValleuValAlaserglylyvalAlaseranpethrylleargarg 160
 Db 1289 CAATAATATGACATCTGTTGATCTGCGTGGTGCAGATTAATCTTATCCGACGA 1348

QY 161 AlaleuGluilleuThrasnalathrgnglyvalserThreleuencysppropoargleu 180
 Db 1349 GCTTCGAAATTTTAAACAAAGCAGACAGTGTGCTTGTGTCTCTCCACAGCTA 1408

QY 181 CysThraspaanglyllewetiellaatrpanglylleuargleuarglaaglyleu 200
 Db 1409 TGACAGTATATGCGATTATGATTGCA-----TGATGTCCTTGGAGTAC 1435

QY 201 G1yleleuHieapllieguglylleargtyrlyuprolvscyeproleuglyvalasp 220
 Db 1436 -----TGATGTCCTTGGAGTAC 1456

RESULT 12
 AK055441 2208 bp mRNA linear PRI 30-JAN-2004
 LOCUS AK055441
 DEFINITION Homo sapiens cDNA FLJ30879 f1s, clone FEBR2004592, highly similar to Homo sapiens mRNA for putative sialoglycoproteinase type 2.
 ACCESSION AK055441
 VERSION AK055441.1 GI:16550166
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Chordata; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
 1 Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H., Sekine,M., Oobayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y., Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M., Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T., Kikawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H., Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yoshida,M., Horita,T., Kusano,J., Kanehori,K., Takahashi-Fuji,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuki,H., Oshima,A., Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsumura,H., Ichihara,T., Shihata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T., Kumegai,A., Takemoto,M., Kawakami,B., Yamazaki,Y., Komiyama,K., Kuniyagi,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M., Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hiro,M., Ohmori,Y., Kawabara,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,R., Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Yagashi,T., Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J., Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K., Nagase,T., Nomura,N., Kikuchi,H., Masuhio,Y., Yamashita,R., Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Iosagi,T. and Sugano,S.
 Complete sequencing and characterization of 21,243 full-length human cDNAs
 Nat. Genet. 35 (1), 40-45 (2003)
 14702039

JOURNAL
 PUBMED
 REFERENCE
 AUTHORS
 Nishi,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,

Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kamura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Kanda, K., Magetsuma, M., Murakawa, K., Kanehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, K., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.,
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2208)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES
Source
Location/Qualifiers
1..2208
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FEBRA204592"
/tissue_type="brain"
/clone_id="FEBRA2"
/dev_stage="fetus"
/note="cloning vector: pME18FL3"
344..1438
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB70923.1"
/db_xref="GI:16550167"
/translation="MLLTITKAGVFFKSRKRYVBFSPNPHPTGLFKHIVGIET
SCDDTAADVDRGVNAGEATHSOTENVHKTGQVPPAAQLHENTORIOBALASAS
GVSPSDLSAIRTTKPGIATSLVGVSDFLVGLQKCPPIPIHMEHAAHLLTRITNK
VEPFLVLLISGHCILLVGVSDFLVGLQKCPPIPIHMEHAAHLLTRITNK
MSGKALIEHLAKQGNRFHFDIKPILHAKNCDPFTGLOHATYDKIIMKKEGEIEKG
QIISAAADIATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVIVVAGSVANFCIR
RAELILNATGCTILCPPRCTDNGIMIA"

ORIGIN
Alignment Scores:
Pred. No.: 3.57e-95 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 9 Gaps: 1

US-10-649-273-2_copy_176_414 (1-239) x AK055441 (1-2208)

QY 1 LeuleuAlaIeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 869 CTGTTGGCATTAGTTCAGAGGAGTTTCAGATTTCTGCTTCTTGGAAAGCTTTGACATCA 928
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLysSerLeuIleLysHisProGlu 40
DB 929 GCACCAAGGTGACAGCTTGACAGGTGCAGAGAGACTTCTTTAATAAATCATCCAGAG 988
QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 989 TGCTCCACCAATGAGTGGTGGGAAGCCATAGACATTGGCCAAACAGAAATAGATT 1048
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 1049 CATTTGACATCAAACTCCCTTCGATCATCTAATAAATTTGATTTCTTTACTGGA 1108
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGlyIleGluLys 100

DB 1109 CTTCAACACGTTACTGATTAATAATATGAAAAAGGAAAAAGAGGATTTAGAGAG 1168
QY 101 GYGINIIIELEUSESERIALAIAAPIIIEAIAIETHVALGIMHISPHMETALACYS 120
DB 1169 GGGCAAAATCTCTCTTCAGACAGACAGATTCGTCACAGTACAGACCAATGGCATGT 1228
QY 121 HISELVAIIYASRTHRHISARGAIAIIELEUPHECYELYSGLINARGASPLEUAPRO 140
DB 1229 CATCTTGCAAAAAGAACATCGGCTATTCTGTTTGTAGCAGAGACATCTGTACCT 1288
QY 141 GINAEAAENALIAVALLEUVALIASERGILYIVALIASERASPHETHRIIEARGATG 160
DB 1289 CAAATTAATGACGTAAGTGGTTCATCTGTTGGTGGAGTAACCTGTATCCACAGA 1348
QY 161 ALALEUGIIIELEUTHRASNAIATRGINCSTHLEULEUCYSPROPROAIGLEU 180
DB 1349 GCTCTGGAATTTTAAACAACCAACACAGTGCATCTTGTGTCTCTCCACACATA 1408
QY 181 CYSTHRASPAANGIYIIELETHLEAIAITPAANGIYILEGLUARGLEURGALIGLYLEU 200
DB 1409 TGCACCTGATTAATGGCATTAATGATTGCA----- 1435
QY 201 GYIIELEUHIASPIIEGLUGIYILEARGTYRGUPLROLYSCYSPROLEUGIYVALASP 220
DB 1436 -----TGANGTCTCTTGGAGTAGAC 1456
QY 221 ILESERLYSGIIVAGIYGLUALASERIIIELYVALPROGLINLEULYSMETGIUIIE 239
DB 1457 AATTCAAAAGAAAGTTGAGAAACCTTCATTAAGTACCAATTAATAATGAGATA 1513

RESULT 13
BC078974 1546 bp mRNA linear ROD 03-AUG-2004
LOCUS BC078974
DEFINITION Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.
ACCESSION BC078974
VERSION BC078974.1 GI:50926879
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 1546)
Strasberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shemmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, D., Hsieh, F.,
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, W.J., Usdin, T.B., Toshiyuki, S.,
Carrinoci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Pahey, O., Helton, B., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Kravinsky, M.I., Skalska, U., Smallus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.,
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL PUMED 12477932
REFERENCE 2 (bases 1 to 1546)
AUTHORS Director MGC Project.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT

Contact: MGC help desk
Email: cgaabs-remail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAX Plate: 182 Row: f Column: 6
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

source

1. 1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/tissue_type="Testis, rat (Brown Norway)"
/clone_id="N1H_MGC_237"
/lab_host="DH10B"
/note="Vector: pExpress1"
144. >1546
/codon_start=1
/product="Unknown (protein for IMAGE:7111906)"
/protein_id="AAH78974.1"
/db_xref="GI:50926880"

CDS

/translation="MLMSKTAGAI PRPNSNVGRIRFNVQPRALFHHKLVGIET
/SCDPTAAVVDENYGEALSHQTEVHLKTGQVPPAQQIHRNRIIVEALSLAS
GVSPSDLSAATTKPGIALSVGLSFDLVNQKQPIPIHMEALALIRLTHK
VGPFVLITISGHCIALVQSGDPLILGSDLPQDMLDKVRRSLIKHPCST
MSGKALEHAKGKRNFRHTIPPMONAKNDFSGVQHTDKLITKEKEGEBK
QIISSAADIAAVOHTACHIAKTRHAILTCQKNLSPANAVALVSGVASNIYIR
RALEIVANATQCTLCPPRLCTNGMILANAGERLISAGAILADVDIDYEPAGE
IVSWLRVLTALTEDIDSVSTHTVAHSPILSGSRGANIQSMSCCTQIVYMETVHTL
NINERKSKKKKKKKK"

ORIGIN

Alignment Scores:

Pred. No.: 8.27e-88 Length: 1546
Score: 983.00 Matches: 188
Percent Similarity: 92.06% Conservative: 9
Best Local Similarity: 87.85% Mismatches: 17
Query Match: 79.27% Indels: 0
DB: 10 Gaps: 0

US-10-649-273-2_copy_176_414 (1-239) x BC078974 (1-1546)

QY 1 leuleuValaValaGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 669 CTGTTGGCGTAGTTCAGAGTGTTCAGATTTCCTGCTCCGCGAAGTCCCTGACATA 728
QY 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleGlyHisProGlu 40
DB 729 GCGCCAGGCGACATGCTTGACAAAGTGGCAAGAACTTTCTTTATCAACATCCAGAA 788
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 789 TGTTCATCATGAGTGTGGGAAAGCTATAGAACTTTGGCCAAAGAAATGATTC 848
QY 61 HisPheAspIleLysPProPheLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 849 CACTTACTACTCATCATCCAGAACTCAGAACTGATTTCTTTTAACTGCGGA 908
QY 81 leuGlnHisValThrAspLysIleIleMetLysGlnGlyLysGlnGlnGlyTlGlnLys 100
DB 909 CTTCACATGTCACCGATAGCTAATTAACACAAAGGAAAAAGAAAGCAATGGAAG 968

QY

101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120

DB

969 GGGCAAAATCTGTCATCAGCCGACGACATTCCTGCTCGGTACAGCAGCAACAGCGTGC 1028

QY

121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140

DB

1029 CACCTTGCGCAAAAGAACATCGTGTCTATTCGTTTGGCAGCAAGAAATTTGCTACTC 1088

QY

141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTYlleArgArg 160

DB

1089 CCAGCTAACGACATATTAGTGTGTGTCGAGGCTTGCAAGTAACCTGTATCATCCGAA 1148

QY

161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysPProPheArgLeu 180

DB

1149 GCATTGAAAATGTACCAATACCAACATGACATTTGTTGTGTCCTCCGAGACTG 1208

QY

181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGlnArgLeuArgAlaGlyLeu 200

DB

1209 TGCACAGACATGATGATCATGATTCATGGAATGGAATGAATTAAGATTACCTGCTG 1268

QY

201 GlyIleLeuHisAspIleGlnGlyIleArgTYlleArgTYlleProLys 214

DB

1269 GCATTATTACATGATGTAGAGACATCCGATACGAACCAAG 1310

RESULT 14

BC051211

BC051211 1109 bp mRNA 1linear ROD 15-APR-2003

LOCUS

Mus musculus, clone IMAGE:1327545, mRNA.

DEFINITION

BC051211

ACCESSION

BC051211.1 GI:29881634

VERSION

BC051211.1 GI:29881634

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Mus musculus (house mouse)

REFERENCE

1 (bases 1 to 1109)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (14-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk
Email: cgaabs-remail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAX Plate: 113 Row: b Column: 1
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

source

Location/Qualifiers
1. 1109
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1327545"
/tissue_type="Thymus gland, mouse"
/clone_id="Soares_thymus_2nbMT"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac"

ORIGIN


```

Qy 161 AlaleuGlulIleuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 574 GAGCTGCAGAGCTCTGCAAAATGCAAAAGCGTTTTCCTTCTGCTCTCTCCAAAGCTG 633
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 634 TGCACCGATTAATGCTGTATGATGTCATGGAATGGCATTGAAAGGTTGCGTCAGATGT 693
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrgIuProIysCysProIeuGlyValAsp 220
Db 694 GGAATTATACAGTACTGATGGCAGCCGCTACGAAACCAAAAGCTCCCTTGGAATTGAT 753
Qy 221 IleserIysGluValGlyGluIleSerIleIysValProGlnLeuIys 236
Db 754 ATTTCAAAAGAGTGAAGAGGATTCCATCAAAGTCCCAAGACTAAG 801

```

Search completed: February 16, 2005, 18:03:04
 Job time : 3816.22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:09:32 ; Search time 455.139 Seconds
(without alignments)
3108.540 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVGVSPFLILGKSLDI.....DISKEVGASIKVPLKMEI 239

Scoring table: BIOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model DEV=xlh
-Q=/cg2_1/USPTO_spool/US10649273/runat_14022005_114702_16389/app_query.fasta_1.1429
-DB=N_Geneseq_16Dec04 -QPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US10649273_QCGN_1.1_1057_@runat_14022005_114702_16389 -NCPu=6 -ICPU=3
-NO_MMAP -LARGEQUERRY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004as:*
- 13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	1240	100.0	1416	ABX70950 Novel hum
2	1240	100.0	1526	ABs76639 DNA encod
3	1240	100.0	2058	ABs93268 Human O-s
4	1240	100.0	2197	ABs76635 DNA encod
5	1240	100.0	2572	ABT23207 Human pro

6	1203	97.0	1820	AAU46856
7	1203	97.0	1820	ACA60887
8	1203	97.0	1821	ABs57020
9	1059	85.4	2208	ADA52832
10	1059	85.4	2208	ADQ24627
11	995.5	80.3	3358	ADs31345
12	870	70.2	1572	ABQ75508
13	599	48.3	2734	AA584622
14	468	37.7	371	ADL86725
15	468	37.7	371	ADL86726
16	332.5	26.8	1601	ABU24633
17	332.5	26.8	3656	ABU24632
18	322	26.0	1557	ACA38454
19	257	20.7	1146	ACA26804
20	248.5	20.0	1000	ABK37804
21	248.5	20.0	1000	ABK37804
22	248.5	20.0	1044	ACA39102
23	248.5	20.0	94750	AA28551
24	247	19.9	936	ACA20445
25	247	19.9	1053	ADL03120
26	241	19.4	1032	ACA43173
27	239	19.3	4360	ABU48239
28	235.5	19.0	1092	ABQ90383
29	234	18.9	300	AAA00934
30	232	18.7	676	AAH08019
31	232	18.7	1385	AAH15110
32	231	18.6	1026	AA554064
33	231	18.6	1026	ACA42146
34	231	18.6	1026	ADG73341
35	231	18.6	1059	ABD02280
36	231	18.6	1206	ABD02197
37	229	18.5	1026	ADG73343
38	224	18.1	1029	AA553309
39	224	18.1	1029	ACA34150
40	224	18.1	11000	AA742063_05
41	217	17.5	9967	ADT05493
42	217	17.5	85814	ADT05644
43	214	17.3	1044	ABU08591
44	213	17.2	1014	AA556045
45	213	17.2	1014	ACA51431

ALIGNMENTS

RESULT 1	
ABX70950	ABX70950 standard; cDNA; 1416 BP.
XX	ABX70950;
AC	ABX70950;
XX	
DT	05-MAR-2003 (first entry)
XX	
DE	Novel human cDNA sequence #175.
XX	
KW	Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW	Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW	insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;
KW	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW	Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW	differentiation; stem cell growth factor; haematopoiesis; chemokine;
KW	haemostatic; antiinflammatory; expressed sequence tag; EST.
OS	Homo sapiens.
XX	
PN	WO200281731-A2;
XX	
PD	17-OCT-2002.
XX	
PF	29-JAN-2002; 2002WO-US001222.

XX 30-JAN-2001; 2001US-00774528.
XX (HYSE-) HYSEQ INC.
XX (GOOD/) GOODRICH R W.
XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F,
PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Dmanac RT;
XX WPI; 2003-058563/05.
XX
XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
PT or lymphoid cell disorders, bone disorders, mechanical and traumatic
PT disorders, coagulation disorders, and inflammatory diseases.
XX
XX Claim 1, Page: 612pp; English.
XX
XX This invention relates to the cDNA sequences encoding an isolated novel
CC human polypeptide. The protein encoded by the nucleic acid of the
CC invention is useful for treating central and peripheral nervous system
CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
CC bacterial, viral or fungal infections; allergic conditions such as
CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
CC disease, anaphylaxis). The protein may be used to inhibit the growth,
CC infection or function of infectious agents such as bacteria, fungi,
CC viruses, or to effect bodily characteristics, biorthylms or circadian
CC cycles of rhythms. The protein may also have
CC proliferation/differentiation, stem cell growth factor, haematopoietic,
CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
CC activities. The cDNA sequences of the invention are useful for expressing
CC recombinant protein for analysis. The present sequence represents a novel
CC human cDNA sequence of the invention, this sequence is an expressed
CC sequence tag (EST) and was identified using subtractive hybridisation
XX
SQ Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 3-21e-132 Length: 1416
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8 Indels: 0
DB: Gaps: 0
US-10-649-273-2_COPY_176_414 (1-239) x ABX70950 (1-1416)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 586 CTGTTGGCATTAAGTTCAGAGATTTCAGATTTCGCTTCCTGGAAGCTTTGACATTA 645
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAspLeuSerLeuIleLysHisProGlu 40
DB 646 GCACCAAGGTGACATCTTGACAGAGGTGGAGAGACCTTCTTATATAAACAATCCAGAG 705
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 706 TGCTCCACCAAGTGTGGTGGGAAACCATAGAACATTTCGCCAACAAGAAATGATTT 765
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 766 CATTTTGACATCAACCTCCCTTCATCATGCTTAATAAATTTGATTTTCTTTACTGGA 825
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluValGlnGlyIleGlnLys 100
|||||

DB 826 CTTCAACACGTTACTGATTAATAATTAATGAAAAAGAAAAAGAGAGATTTAGCAAG 885
QY 101 GLYGlnIleLeuSerSerIleAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 886 GGGCAAAATCTCTCTTCAGACAGACATGTGCTGCACAGTACACAGACAATGCGATGT 945
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 946 CATCTTGAAAAGAAACACATCGGCTATTCTGTTTGTAGCAGAGACATCTTTACTCT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnAspThrIleArgArg 160
DB 1006 CAAAATTAATGACGTAAGTCTGTCATCTGCTGGTGTCGAAGTAACCTTATATCCGACA 1065
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnGlyThrLeuLeuCysProProProAlaGlu 180
DB 1066 GCTCTGGAATTTTAAACAACCAACACAGTCACTTTGTGTCTCTCCACACATTA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1126 TGCACGTATATAGCATTTATGATTCATGAAATGATGTAAGACTACGTGCTGCTTG 1185
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1186 GGCATTTTACATGACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGAGATGAC 1245
QY 221 IleSerLysGluValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1246 AATATCAAAAGAAAGTTGGAGAACCTTCATTAACCAATTAATAATGAGATA 1302
RESULT 2
ABS76639
ID ABS76639 standard; DNA; 1526 BP.
XX
AC ABS76639;
XX
DT 11-DEC-2002 (first entry)
XX
XX DNA encoding novel human metalloprotease MP1 fragment #1.
DE
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder; gene; db.
XX
OS Homo sapiens.
XX
PN MO200272751-A2.
XX
PD 19-SEP-2002.
XX
PF 05-FEB-2002; 2002WO-US003353.
XX
PR 05-FEB-2001; 2001US-0266518P.
XX
PR 10-APR-2001; 2001US-0282814P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Chen J, Feder J, Nelson TC, Duclos F, Krysiek S;
XX WPI; 2002-723329/78.
XX
DR P-PSDB; ABS96487.
XX
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.

XX
PS - Disclosure; Page 462-463; 473pp; English.
vz

The invention describes an isolated nucleic acid molecule (I) encoding metalloprotease (MP-1). (I) is useful for preventing, treating, or ameliorating a medical condition, particularly an immune disorder, an aberrant glutamate transport or motor neuron disorder, such as amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like condition. The compositions and methods are also useful for diagnosing, prognosticating, treating, ameliorating and/or treating disorders associated with MP-1 activity, e.g. diabetes, cancer, reproductive disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome, or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease, Parkinson's disease, Huntington's disease or Tourette syndrome), liver and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic fibrosis) and vascular, inflammatory and neurological disorders (e.g. Alzheimer's disease or Parkinson's disease). This sequence represents a metalloprotease MPI polynucleotide

SQ Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3,57e-132	Length:	152
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x ABS76639 (1-1526

QY	1	LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspIle	20
Db	85	CTGTGGCATTAAGTTCAAGAGATTTCAAGATTTCTGCTTTGGAAAGCTTTTGGACATA	14
QY	21	AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysIleSProGlu	40
Db	145	GCACCAAGGTGACATGCTTGACAAAGGTGGCAAGAAAGACTTTCTTTAATTAACATCCAGAG	20
QY	41	CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe	60
Db	205	TGCTCCACACATGAGTGTTGGGAAAGCCATATGAAACATTTGGCCAAACAAAGAAATATGATTT	26
QY	61	HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	80
Db	265	CATTTTGACATCAAACTCCCTTGACATGCTGTAATATGTGATTTTCTTTTACTGGA	32
QY	81	LeuGlnHisValThrAspLysIleIleMetLysGlnLysGlnGlnGlyIleGluLys	100
Db	325	CTTCAACACGTTACGTATTAATTAATTAAGAAAAGAAAGAAAGGATATTGAGAG	38
QY	101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys	120
Db	385	GGGCAAACTCTGTCCTTCAGCACAGACATTTGCTCCACAGTACAGCACAAATGGCATGT	44
QY	121	HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro	140
Db	445	CATCTTGGAAGAAAGAACACATGGGGCTATTCGTTTGTTAAGCAAGAGACTTGTAACCT	50
QY	141	GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg	160
Db	505	CAAAATTAATGACGATACGTGTTGCATCTGCTGCTGCGCAAGTAACCTCTAATACCGAGA	56
QY	161	AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	180
Db	565	GCTCGGAAATTTTAAACAACCAACACAGTGCACTTGTGTGTGCTCTCCCAAGACTA	62
QY	161	CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuAspAlaGlyLeu	200
Db	625	TGCATGATTAATAGGACATTATGATGTGATGAAATGATTAAGAAACATACGTGCTGGCTTG	68

201 G L y l e l e n H i s A s p I l l e g l u g l y l e a r g T y r g l u P r o l y s C y s P r o l e n g i y V a l a s p 220
 685 G G C A T T T T A C A T G A C A T A G A G C A T C C G C A T T A G A A C C A A A A T G T C C T C T T G A G A G T A G A C 744

Dy 221 ILeSerLygsluValGlYgluaIAserIlElyVAlProGlnEuLysMetGluIle 239
|||
745 ATATCAAAAGAAAGTTGGAGAAGCTTCATAAAGTACACAATTAATAATGAGATA 801

RESULT 3

ID ABA93268 standard; cDNA; 2058 BP.

AC ABA93268;

DT 19-APR-2002 (first entry)

DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.

KW Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.

Homo sapiens.

FH	Key	Location/Qualifiers
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

PT / *tag= a

2000

XX

.....

✕ ✕

1

XX

XX

DR P-PSDB; ABB05481.

PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful

XX

XX

CC (OSGPLP). The present invention also describes: (1) the preparation of

CC prevention and/or treatment of related diseases; (4) utilising the OSG

CC an antibody against the OSGPLP protein; and (5) the use of

CC and antibodies in treating diseases related to the abnormal

CC	and in preparing the medicine composite for the treat
XX	
SQ	Sequence 2058 BP, 637 A; 400 C; 410 G; 611 T; 0 U; 0
Alignment Scores:	
Pred. No.:	5,45e-132 Length: 2058E

Percent Similarity: 100.00% Conservative: 0

Dy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 635 CTGTGGCATTAGTCAAGAGTTTCAGATTTCTCGCTTTCGAAGCTTTGCACATA 69

```

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleuShisProGlu 40
DB 695 GCACCAAGTGAATGCTTGAACAGGTGCGACAGAGACTTTCTTTAATAAATCAGAG 754
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 755 TGCTCCACATGAGTGGTGGGAAAGCCATAGAACATTTGGCCAAACAGGAATATGATTT 814
QY 61 HisPheAspIleuLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 815 CATTGTGACATCAAACTCCCTGATCATGCTAAATAATGGAATTTCTTTTACATGGA 874
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGlyIleGluLys 100
DB 875 CTTCAACACGTTACTGATTAATTAATTAAGAAAAGAAAAGAGAGAGATTTTGAAGAG 934
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 935 GGGCAAAATCCTGTCTTGACAGACAGACATTCCTGCACAGTACAGACACAAATGGCATGT 994
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 995 CATCTTGGAAAAGAACACATCGGGCTATTCCTGTTTGAAGCAGAGAGACTTGTTACCT 1054
QY 141 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
DB 1055 CAAAATAATGCAATGCTGGTTCATCTGATGCTGATGCTGCAAGTACTTCTAATCCGCGAA 1114
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1115 GCTCTGGAAATTTTAACAAGCAACAGCAACAGTCTTTGTGTGCTCTCTCCAGACATA 1174
QY 181 CysThrAspAsnArgIleIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1175 TGCACTGATTAATGCAATATGATTCATGCAATGATGTAAGCACTACGTCCTGCTTG 1234
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgGlyProLeuGlyValAsp 220
DB 1235 GGGATTTTAACATGACATGAAAGGCAATCCGCTATGAAACAAAATGTCCTTTGAGATGAC 1294
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1295 ATATCAAAAGAGATTGGAGAAAGCTTCCATTAAGATCCATTAATTAATTAATGAGATTA 1351

```

RESULT 4

ABST76635
ID ABST76635 standard; DNA; 2197 BP.

XX AC ABST76635;

DT 11-DEC-2002 (first entry)

DE DNA encoding novel human metalloprotease MPI.

XX
KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
genital wart; metabolic disorder; premature puberty; Kallman syndrome;
Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
liver disease; renal disease; immune disorder; rheumatoid arthritis;
acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
neurological disorder; gene; ds.

OS Homo sapiens.

PN W0200272751-A2.

PD 19-SEP-2002.

PF 05-FEB-2002; 2002WO-US003353.

XX

PR 05-FEB-2001; 2001US-0266518P.
PR 10-APR-2001; 2001US-0282814P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PA
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
PI
XX MPI: 2002-723329/78.
XX P-PsDB; ABG96478.
DR

PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.

PS Claim 1; Fig 1A-C; 473pp; English.

CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
CC ameliorating a medical condition, particularly an immune disorder, an
CC aberrant glutamate transport or motor neuron disorder, such as
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC condition. The compositions and methods are also useful for diagnosing,
CC prognosticating, treating, ameliorating and/or treating disorders
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC Alzheimer's disease or Parkinson's disease). This sequence represents a
CC metalloprotease MPI polynucleotide
XX

Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,98e-132	Length:	2197
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x ABST76635 (1-2197)

```

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 756 CTGTTGGCAATGATGTTCAAGAGATTTTCAGATTTCTTCTTGGAAGCTTTGGACATA 815
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleuShisProGlu 40
DB 816 GCACCAAGTGAATGCTTGAACAGGTGCGACAGAGACTTTCTTTAATAAATCAGAG 875
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 876 TCCTCCACATGATGATGTTGGGAAAGCCATAGAACATTTGGCCAAACAGGAATTAATTT 935
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 936 CATTGTGACATCAAACTCCCTGATCATGCTAAATAATGGAATTTCTTTTACTGGA 995
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGlyIleGluLys 100
DB 996 CTTCAACACGTTACTGATTAATTAATTAAGAAAAGAAAAGAGAGATTTTGAAGAG 1055
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1056 GGGCAAAATCCTGTCTTGACAGACAGACATTCCTGCACAGTACAGACACAAATGGCATGT 1115
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

```


QY 161 AAlaLeuGluIleuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 1149 GCTCTGGAATTTTAAACAAAGCAACAGTGCCTTTGTTGTGCTCTCCAGACTA 1208
QY 181 CysThrAspAnglyIleMetIleAlaTrpAnglyIleGluArgLeuArgAlaGlyLeu 200
Db 1209 TGCACGTATATAGCATTAATATATGCAATGGAATGATTTGAAGAAGCTACGTGCTGGCTTG 1268
QY 201 GAlIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db 1269 GGCATTTTACATGACATAGAAAGCATCCGCTATGAAACCAAAATGCTCTTGAGATGAC 1328
QY 221 ILeSerIyGluValGlyGluAlaSerIleIyValProGlnLeuIyMetGlyIle 239
Db 1329 ATATCAAAAGAGTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1385
RESULT 6
AAD46856
XX ID AAD46856 standard; cDNA; 1820 BP.
XX
XX AAD46856;
AC
DT 27-JAN-2003 (first entry)
XX
DE Human glycoprotease 28472 cDNA.
XX
KW Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
KW 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;
KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;
KW hypertension; ischaemic heart disease; obesity; myocardial infarction;
KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;
KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;
KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;
KW chromosome mapping; tissue typing; gene therapy; neuroprotective;
KW cytoskeletal; anorectic; cardiac; haemostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 146..1390
FT /*tag= a
FT /product= "Human 28472 protein"
FT /note= "This region is specifically claimed as SEQ ID NO:
FT 6 in claim 1 of the specification"
XX
XX MO200274960-A2.
XX
XX 26-SEP-2002.
XX
XX 08-NOV-2001; 2001MO-US051427.
XX
XX 08-NOV-2000; 2000US-0246768P.
XX 08-NOV-2000; 2000US-0246772P.
XX 15-NOV-2000; 2000US-0249185P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Ielby KR, Kapeller-Libermann R, Gluckmann M;
PI WPI; 2002-75998/82.
XX
XX P-PDB; AAE29234.
XX
XX New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
XX useful for diagnosing and treating cancer, immune, cardiovascular,
XX hematopoietic, brain, pain, metabolic, liver or platelet disorders, and
XX in pharmacogenomics.
XX
XX Claim 1, Fig 8, 178pp; English.
XX
XX The present invention relates to novel 38650, 28472, 5495, 65507, 81588
XX or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
XX protease or seven transmembrane domain (7TM) receptor family members.
XX Sequences of the invention are useful in diagnosing and treating cancer

CC or aberrant cellular proliferation and/or differentiation (e.g. colon or
CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
CC myocardial infarction, thrombus) including endothelial cell disorders
CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet
CC disorders. They are also useful in screening assays, predictive medicine
CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
CC and pharmacogenetics) and prophylactic and therapeutic methods. The
CC nucleic acids may also be used in chromosome mapping, tissue typing and
CC forensic biology and as surrogate markers. Sequences of the invention are
CC also used in gene therapy. The present sequence is human glycoprotease
XX 28472 cDNA
XX
SQ Sequence 1820 BP, 543 A, 365 C, 393 G, 518 T, 0 U, 1 Other;
XX
Alignment Scores:
Pred. No.: 8-226-128 Length: 1820
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
DB: Gaps: 0
US-10-649-273-2_COPY_176_414 (1-239) x AAD46856 (1-1820)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIySerLeuAspIle 20
Db 671 CTTTGGCATTAATGTTCAAGAGATTTTCAAGATTTTCTGCTTGGAAAGCTTTGACATA 720
QY 21 AlaProGlyAspMetLeuAspIyValAlaAlaGluLeuSerLeuIleIyHisProGlu 40
Db 721 GCACCAAGTGAACATCTTGAACAGGTGGCAAGAGCTTTCTTAATAAATCACTCAGAG 790
QY 41 CysSerThrMetSerGlyGlyIyValAlaIleGlnHisLeuAlaIyGlnIyAsnAAspPhe 60
Db 791 TCTCCACCATAGAGTGGTGGAAAGCATAGAACATTTGGCCAAACAGAAATAGATTT 850
QY 61 HisPheAspIleIyAspProProLeuHisIleAlaIyAsnCysAspPheSerPheThrGly 80
Db 851 CATTTTGACATCAAACTCCCTGTCATCATGTCTTAATAATTTGATTTCTTTTACTCGA 910
QY 81 LeuGlnHisValThrAspIyIleIleMetIyIyGluIyGluGlyIleGluIyIy 100
Db 911 CTTCAACACGTTATCGATTAATAATATAGAAACAGGAAACAGAGAGGATTTGAGAG 970
QY 101 GAlIyGlnIleLeuSerSerIleAlaAspIleAlaIleValGlnHisIleThrMetAlaCys 120
Db 971 GGGCAAAATCTGTTCTTCAGACAGACATGCTGCCACAGTACAGCACCAATGGCATGT 1030
QY 121 HisLeuValIyAspArgThrHisArgAlaIleLeuPheCysIyGlnArgAspLeuLeuPro 140
Db 1031 CATCTTGCAAAAGAACACATCGGCTATCTCTTTTGAACAGAGACATCTGTAACT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyIyValAlaIleSerAspPheIyIleArgArg 160
Db 1091 CAAATAATATGACATGATCGTTCATCTGGTGTGTCGCAAGTAACTTATATCCGACAG 1150
QY 161 AAlaLeuGluIleuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 1151 GCTCTGGAATTTTAAACAAAGCAACAGTGCCTTTGTTGTGCTCTCCAGACTA 1210
QY 181 CysThrAspAnglyIleMetIleAlaTrpAnglyIleGluArgLeuArgAlaGlyLeu 200
Db 1211 TGCACGTATATAGCATTAATATATGCAATGGAATGATTTGAAGAAGCTACGTGCTGGCTTG 1270
QY 201 GAlIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db 1271 GGCATTTTACATGACATAGAAAGCATCCGCTATGAAACCAAAATGCTCTTGAGATGAC 1330
QY 221 ILeSerIyGluValGlyGluAlaSerIleIyValProGlnLeuIyMetGlyIle 239

DB 1331. ATATCAAAAGAGTTGGAGAGCTTCCATAAAAGTACCAATTTAAAATGAGATA 1387
 RESULT 7
 ID ACA60887 standard; cDNA; 1820 BP.
 AC ACA60887;
 XX
 DT 08-JUL-2003 (first entry)
 XX
 DE Human cDNA 28472 encoding a glycoprotease.
 XX
 KM Human; 88; gene; cancer; aberrant cellular proliferation;
 KM differentiation; immune disorders; heart disorder; brain disorder;
 KM cardiovascular disorder; endothelial cell disorder; pain disorder;
 KM hematopoietic disorder; blood vessel disorder; metabolic disorder;
 KM liver disorder; platelet disorder; glycoprotease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 146..1390
 FT /tag= a
 FT /product= "Glycoprotease"
 FT /note= "This CDS is specifically claimed in claim 1"
 XX
 PN US2003009017-A1.
 XX
 PD 09-JAN-2003.
 XX
 PF 08-NOV-2001; 2001US-00012140.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (LEIBY K R.
 PA (KAPELLE L-BERMAN R.
 PA (GLUC/) GLUCKSMANN M A.
 PI
 PI Leiby KR, Kapeller-Libermann R, Gluckemann MA;
 PI WPI; 2003-428888/40.
 DR P-PSDB; AB009369.
 XX
 XX
 PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic
 PT and liver disorders.
 PT
 PS Claim 2; Fig 8; 90pp; English.
 XX
 XX
 CC The invention relates to an isolated 38650 (encoding adenosine
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
 CC a sequence which is at least 60% identical to the six nucleic acids or
 CC their open reading frames, fragments of at least 15 nucleotides,
 CC naturally occurring variants, or a DNA insert of the plasmid deposited
 CC with the American Type Culture Collection as Accession No. not defined in
 CC the specification, which encodes the amino acid sequence). Also included
 CC are a host cell containing the nucleic acids (used to produce the
 CC proteins), the encoded proteins, an antibody that selectively binds to
 CC the polypeptide, and identifying a compound that binds to/modulates the
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
 CC methods are useful for diagnosing, treating cancer, aberrant cellular
 CC proliferation and/or differentiation, immune disorders, heart disorders,
 CC cardiovascular disorders including endothelial cell disorders,
 CC hematopoietic disorders, blood vessel disorders, brain disorders, pain
 CC and metabolic disorders, liver disorders and platelet disorders (many
 CC examples of these disorders are given in the specification). The present
 CC sequence is the Human cDNA 28472 encoding a glycoprotease
 CC

SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
 Alignment Scores:
 Pred. No.: 8 22e-128 Length: 1820
 Score: 1203.00 Matches: 232
 Percent Similarity: 98.33% Conservative: 3
 Best Local Similarity: 97.07% Mismatches: 4
 Query Match: 97.02% Indels: 0
 DB: 8 Gaps: 0
 US-10-649-273-2_COPY_176_414 (1-239) x ACA60887 (1-1820)
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyGlySerLeuAspPle 20
 DB 671 CTTGGGCAATTGTTCAAGAGATTTCAGATTTTCCTCTTGGAAGCTTTGGACATA 730
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
 DB 731 GCACCAAGTGAATGATGTTGACAAGGTGGCAGAGACTTTCTTAATAAACATCCAGAG 790
 QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
 DB 791 TGTCTCCACCATGATGATGTTGGAAAGCCATAGAACATTTG3CCAAACAGAAATAGATT 850
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
 DB 851 CATTTTGACATCAACCTCCCTTCATCATGCTTAATAATTTGATTTTCTTTACTGGA 910
 QY 81 LeuGlnHisValThrAspLysIleIleMetLysGluLysGluGlyIleGluLys 100
 DB 911 CTTCAACAGCTTACTGATTAATAATATGAAAACAGAAAACAGAGAGATTTGAGAA 970
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
 DB 971 GGGCAATCTCTCTTCAGACAGACAGCATTTGCTCCACAGTACAGACACAATGGCATGT 1030
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
 DB 1031 CATCTTGAAAAGAACACATGGGCTATCTGTTTGTGAAGCAGAGACTGTATACCT 1090
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
 DB 1091 CAATTAATATGCAATGCTGCTTCATCTGCTGCTGCTGCAAGTACTTCTATATCCGACAA 1150
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
 DB 1151 GCTTGCAAAATTTTAACAAACCAACACAGTGCACATTTGTGTGCTCTCCACAGACTA 1210
 QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
 DB 1211 TGCACCTGATTAATGGAATTAATGATTCATGGAATGTGTGAAGACTAGTGTGCTGTG 1270
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgIleArgIleProLysCysProLeuGlyValAsp 220
 DB 1271 GGCAATTTTATCATGACATTAAGAGCATCCGCTATGAACCAAAATGCTCTTGAGATGAGC 1330
 QY 221 IleSerLysGluValGlyGluLysSerIleLysValProGlnLeuLysMetGluIle 239
 DB 1331 ATATCAAAAGAGTTGGAGAACTTCCATRAAAGTACCAATTTAAAATGAGATA 1387
 RESULT 8
 ABS57020
 ID ABS57020 standard; cDNA; 1821 BP.
 AC ABS57020;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE cDNA encoding novel human glycoprotease 28472.
 XX
 KM Cancer; aberrant cell proliferation; aberrant cell differentiation;
 KM breast cancer; ovarian cancer; prostate cancer; colon cancer;
 KM lung cancer; immune disorder; heart disorder; cardiovascular disorder;
 KM

KM endothelial disorder; hematopoietic disorder; blood vessel disorder;
 KM brain disorder; pain; metabolic disorder; liver disorder; diabetes;
 KM platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;
 KM autoimmune disorder; hypertension; atherosclerosis; heart failure;
 KM myocardial infarction; ischemic heart disease; Crohn's disease;
 KM Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
 KM cerebral ischemia; peripheral neuropathy; Alzheimer's disease;
 KM Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 147..1391
 FT /*tag= a
 FT /product= "Glycoprotease 28472"
 FT /note= "Specifically claimed in claim 1"
 FT
 PN WO200277233-A2.
 PD 03-OCT-2002.
 XX
 PD 08-NOV-2001; 2001WO-US046724.
 XX
 PF 08-NOV-2000; 2000US-0246768P.
 XX
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
 XX
 XX WPI: 2003-029938/02.
 DR P-PSDB; ABG71162.
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
 PT hypertension.
 XX
 PS Claim 2, Fig 8A-B; 178pp; English.
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
 CC sequences that encode a human seven transmembrane domain (7TM). The
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
 CC sequences are useful for diagnosing, preventing or treating a subject
 CC with or at risk of developing a disorder, e.g. cancer or aberrant
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
 CC prostate, colon or lung cancer), immune disorders, heart disorders,
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,
 CC liver disorders or platelet disorders. These disorders include carcinoma,
 CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,
 CC ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
 CC cachexia or diabetes. This sequence encodes the novel human glycoprotease
 CC 28472
 XX
 SQ Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;
 XX
 Alignment Scores:
 Pred. No.: 8,22e-128 Length: 1821
 Score: 1203.00 Matches: 232
 Percent Similarity: 98.33% Conservative: 3
 Best Local Similarity: 97.07% Mismatches: 4
 Query Match: 97.02% Indels: 0
 DB: 10 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x ABS57020 (1-1821)
 QY 1 LeuAuaAlaLeuValGlnGlyValSerAspPheLeuLeuGlySerLeuAspIle 20
 DB 672 CCGTGGCACTTGGTCAAGAGATTCAAGATTTCCTCTTGGAAAGCTTTGGACATA 731
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisProGlu 40
 DB 732 GCACCAAGGATGATCTTGTACAAAGGTGCAAGAGACTTCTTTATATAAATCCAGAG 791
 QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
 DB 792 TCTCCACCATGATAGGTGGAGAAAGCATATGGCCAAACAGAGAAATATGATTT 851
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
 DB 852 CATTGTGACATCAAACTCTCCCTGCAATGCTAAATAATGTGATTTCTTTACTCGA 911
 QY 81 LeuGlnHisValThrAspLysIleIleMetLysGluLysGluGlyIleGluLys 100
 DB 912 CTTCAACACGTTACTGATATAAATAATGAACAGAAACAGAGAGATTTAGAGAG 971
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
 DB 972 GGGCAAAATCTGTCTTACAGCAGACATTTGCTGCACAGTACAGCACAATGGCATGT 1031
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgPheLeuPro 140
 DB 1032 CATCTTGAAAAAGAACACATGGGCTATCTGTTTGTAGACAGAGACTGTATACCT 1091
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaLysAsnPheTyrlleArgArg 160
 DB 1092 CAAAATATATGACGACTGTGTGTGATCTGTGTGTGTGCAAGTACTTCTATATCCGACA 1151
 QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
 DB 1152 GCTCTGAAAATTTTAAACAACCAACACAGTGCATTTGTGTGCTCTCCACAGCTA 1211
 QY 181 CysThrAspAsnGlyTlleMetIleAlaTTPaenGlyTlleGluArgLeuArgAlaGlyLeu 200
 DB 1212 TGCACGTATATATGCAATATATGATTTGATGATGATGATGATGATGATGATGATGATG 1271
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleArgTyrlleCysProLeuGlyValAsp 220
 DB 1272 GGCATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1331
 QY 221 IleserLysGluValGlyGluLaserIleLysValProGlnLeuLysMetGluIle 239
 DB 1332 ATATCAAAAGAAAGTTGGAAGAACTTCCATTAAGTACCAATTAATAAATGAGATA 1388
 RESULT 9
 ADA52832
 ID ADA52832 standard; cDNA; 2208 BP.
 XX
 AC ADA52832;
 XX
 XX 20-NOV-2003 (first entry)
 DT Human coding sequence, SEQ ID 400.
 XX
 KW Cytostatic; Anti-inflammatory; Osteoprotic; Neuroprotective; Nocotropic;
 KW Gene therapy; human; secretory protein; membrane proteins; cancer;
 KW Inflammatory disease; osteoporosis; neurological disease; gene; ss.
 OS Homo sapiens.
 XX
 XX Homo sapiens.
 XX
 XX EP1293569-A2.
 XX
 PD 19-MAR-2003.
 XX
 PF 21-MAR-2002; 2002EP-00006586.
 XX
 XX 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isega T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX WPI: 2003-395539/38.
 DR P-PSDB; ADA54471.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 PS Claim 1; SEQ ID NO 400; 205bp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 XX
 SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 3.88e-111 Length: 2208
 Score: 1059.00 Matches: 211
 Percent Similarity: 88.28% Conservative: 0
 Best Local Similarity: 88.28% Mismatches: 4
 Query Match: 85.40% Indels: 24
 Gaps: 1
 US-10-649-273-2_COPY_176_414 (1-239) x ADA52832 (1-2208)
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspLe 20
 Db CTGTTGGCATTAAGTTCAGAGAGATTTCAGATTTTCGCTTTCGAAAGCTTTGACATTA 928
 QY 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleValHisProGlu 40
 Db GCACCAAGGTGACATGCTTGCACAGGTGCAAGAGACCTCTTATATAAACATCCAGAG 988
 QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyAsnArgPhe 60
 Db TGCTCCACCAATGAGTGTGGAGAAAGCCATAGAACATTTGGCCAAACAGGAATGATTT 1048
 QY 61 HisPheAspIleLeuProProLeuHisHisAlaValAsnGlyAspPheSerPheThrGly 80
 Db CATTTTGCATCAAACTCCCTTCGATCATGCTAAATAATGATTTCTCTTAACTGGA 1108
 QY 81 LeuGlnHisValIleThrAspLeuValIleMetCysGlyGluValGluGluValIleGluVal 100
 Db CTTCAACACGTTACTGATTAATAATAATAGAAAAAGAAAAAGAGATTTAGAGAG 1168
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
 Db GGCGAAATTCCTGCTTCGACAGACATGCTCTGCCACAGTACACACAAATGCGCATGT 1228
 QY 121 HisLeuValIleValAspGlyThrHisArgAlaIleLeuPheCysGlyGlnArgAspLeuLeuPro 140
 Db CATCTTGTGAATAACACACATGCGGCTATTCGTTTTTAAAGACAGAGACTGTACTCT 1288
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValValAlaSerAsnPheThrIleArgArg 160
 Db CAAAATAATGACAGTACGCTTGCATCTGCTGCTGCTGCGAAGTAACTTCTGATCCGCA 1348
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
 Db GCTCTGGAATTTTAAACAAAGCAACACAGTGCATCTTGCTGTGTGCTCTCCCGACGACTA 1408
 QY 181 CysThrAspAsnGlyIleMetIleAlaIlePheAsnGlyIleGluValGluLeuArgAlaGlyLeu 200

Db 1409 TGCACGTGATTAATGCGCATTAATGATTCGA----- 1435
 QY 201 GlyIleLeuHisAspIleGluGluGlyIleArgTyrGluProIleCysProLeuGlyValAsp 220
 Db 1436 -----TGATGCTCTCTTGGAGTAGAC 1456
 QY 221 ILeSerIleGlyValGlyGlnAlaSerIleValValProGlnLeuIleValMetGluIle 239
 Db 1457 ATATCAAAAGAGCTTGAGAGAGCTTCATTAAGTACACATTTAAATAATGAGATTA 1513
 RESULT 10
 ADQ24627
 ID ADQ24627 standard; DNA; 2890 BP.
 XX
 AC ADQ24627;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.
 XX
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2004048938-A2.
 XX
 PD 10-JUN-2004.
 XX
 PF 26-NOV-2003; 2003WO-US038193.
 XX
 PR 26-NOV-2002; 2002US-0429739P.
 XX
 PA (PROT-) PROTEIN DESIGN LABS INC.
 XX
 PI Aziz N, Ginsburg WM, Zlotnick A;
 XX WPI: 2004-441208/41.
 XX
 PT Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX
 PS Example 2; SEQ ID NO 7447; 210bp; English.
 XX
 CC The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;
 XX
 Alignment Scores:
 Pred. No.: 5.67e-111 Length: 2890
 Score: 1059.00 Matches: 211
 Percent Similarity: 88.28% Conservative: 0
 Best Local Similarity: 88.28% Mismatches: 4
 Query Match: 85.40% Indels: 24
 Gaps: 1
 US-10-649-273-2_COPY_176_414 (1-239) x ADQ24627 (1-2890)

```

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 1526 CTGTGGCATTAAGTTCAGAGGATTTCAGATTTCCTTGGAAAAGCTTGGACATA 1585
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgGlyLeuSerLeuIleLysHISProGlu 40
DB 1586 GCACGAGGTGACATGCTGACAGGTGCAAGAGAACATCTCTTATATAAACATCCAGAG 1645
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaValGlnGlyAsnArgPhe 60
DB 1646 TGCTCCACATAGTGTGGAGAGCCATAGAACCTTGGCCAAACAGAAATAGATT 1705
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 1706 CATTTGACATCAAACTCCCTTGATCATGTAAATTTGGATTTCCTTTTACCTGA 1765
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlyGluGlyIleGluLys 100
DB 1766 CTTCAACACCTTACCTGATTAATAATATATGAAAAAGAAAAAGAGTATTGAGAAAG 1825
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1826 GGGCAAAATCCGTCTTCAGCAGACAGATGCTGCCACAGATACACACCAATGGCATGT 1885
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1886 CATCTTGGAAGAAACACATCCGGCTATTCGTTTGTAGACAGAGACTGTTACT 1945
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
DB 1946 CAAATATATGACATGCTGTCATCTGTGTGTGTCGCAAGTACCTTGATTCGCGAGA 2005
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 2006 GCTCTGGAATTTTAAACAGCAACAGCAGACATCTTGTGTGTCTCTCCAGACTA 2065
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 2066 TGCACTGATTAATGGCATTAATATGCA----- 2092
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrgLupProLysCysProLeuGlyValAsp 220
DB 2093 -----TGATGTCTCTTGGAGTAGAC 2113
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 2114 ATATCAAAAGAGTGGAGAGCTTCCATPAAAGTACCAATTAATAAATGAGATA 2170

```

RESULT 11
ADE31345/c
ADE31345 standard; DNA; 3358 BP.

```

AC ADE31345;
XX
XX
XX 29-JAN-2004 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide (dthp), SEQ ID NO 100.
XX
XX diagnostic and therapeutic polynucleotide; dthp; antiarteriosclerotic;
XX antiinflammatory; cerebroprotective; antidiabetic; antidiabetic;
XX immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;
XX osteopathic; antiarthritic; antineumatic; cyostatic; hepatotropic;
XX vitamin; haemostatic; anti-HIV; antithyroid; thyromimetic;
XX dermatologic; antibacterial; fungicide; antiparasitic; anticonvulsant;
XX thrombolytic; anticoagulant; anorectic; vasodilator; antidiabetic;
XX gene therapy; protein replacement therapy; human; gene; ds.
XX
XX Homo sapiens.
XX
XX WO2003062376-A2.
XX
XX 31-JUL-2003.
XX

```

```

PR 13-JAN-2003; 2003WO-US001096.
XX
XX 16-JAN-2002; 2002US-0349384P.
PR 17-JAN-2002; 2002US-0349413P.
PR 17-JAN-2002; 2002US-0349946P.
XX
XX (INCY-) INCYTE GENOMICS INC.
PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JR,
PI Yu JY, Tanson O, Yap PB, Amesey SR, Dam TC, Liu TP, Gerslin EH,
PI Perilla CH, Lewis SA, Chen A, Marwaha R, Lan RX, Uraehka ME,
PI Kristnam SR, Kolluru V, Panesar IS;
DR WPI: 2003-636732/60.
XX P-PSDB; ADE31156.
XX
XX New human diagnostic and therapeutic polynucleotides and polypeptides,
PT useful for diagnosing, treating or preventing e.g. leukemia, brain
PT cancer, arteriosclerosis, AIDS, thyroiditis, infections, obesity, stroke
PT or Alzheimer's.
PS Claim 1; SEQ ID NO 100; 634bp; English.
XX
XX The invention relates to a novel isolated human diagnostic and
CC therapeutic polynucleotide (designated dthp). The novel dthp
CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
CC base pairs fully defined in the specification; a polynucleotide
CC comprising a naturally occurring polynucleotide sequence at least 90%
CC identical to the dthp polynucleotide; a polynucleotide complementary to
CC the dthp polynucleotide or its polynucleotide which is at least 90%
CC identical; or an RNA equivalent of any of the polynucleotides mentioned
CC above. The dthp polynucleotides have the following activities:
CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antidiabetic,
CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
CC tranquilizer, osteopathic, antiarthritic, antineumatic, cyostatic,
CC hepatotropic, vitamin, haemostatic, anti-HIV, antithyroid, thyromimetic,
CC dermatologic, antibacterial, fungicide, antiparasitic, anticonvulsant,
CC thrombolytic, anticoagulant, anorectic, vasodilator, and antidiabetic. The
CC novel dthp polynucleotides polypeptide can be used in gene therapy and
CC protein replacement therapy. The dthp polynucleotides or dthp
CC polypeptides are useful for diagnosing, preventing or treating diseases
CC associated with the expression of human molecules. In particular, these
CC diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain
CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung
CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,
CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
CC thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
CC viral, bacterial, fungal or parasitic infection), developmental disorders
CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
CC diabetes, hypopituitarism, hypogonadism, gigantism, goiter) metabolic
CC disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,
CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic
CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),
CC transport disorders (e.g. akinesia or multidrug resistance), or
CC connective tissue disorders (e.g. Paget's disease or rickets). This
CC polynucleotide sequence represents one of the human dthp DNA sequences
CC of the invention.
XX
XX Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;
SQ

```

Alignment Scores:

```

Pred. No.: 1,4e-103 Length: 3358
Score: 995.50 Matches: 209
Percent Similarity: 87.82% Conservative: 0
Best Local Similarity: 87.82% Mismatches: 2
Query Match: 80.28% Indels: 29
DB: 10 Gaps: 1

```

```

US-10-649-273-2_copy_176_414 (1-239) x ADEB31345 (1-3358)
QY      29 ValAlaArgArgLeuSerLeuIleValHisProGluCysSerThrMetSerGly-GlyIy 48
Db      2104 GTGGCAGAGAGACTTTCTTTAATTAATCAATCAAGAGCTCACCACGATGAGTGGGGAA 2045
QY      48 salaiiegluHileuValysGlnGlyAsnArgPheHisPheAspIleYsProProle 68
Db      2044 AGCCATAGAACATTGGCCAAACAGAGAAATGATTTCAATTTGACATCAAAACCTCCCTT 1985
QY      68 uHISHsAlaIleValAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysI 88
Db      1984 GCATCATGCTTAAATAATGTGATTTCTTTACTGACCTTCAACACGTTACTGATAAAT 1925
QY      88 eileMetLysLeuGluysGlnGlu----- 96
Db      1924 AATTAATGAAAAAGGAAAAAGAGAGATATTTCTAATTAGTAAAGTTGAACAGATAAA 1865
QY      97 -----GlyIleGluLysGlnGlnI 103
Db      1864 TATTCTGATGTGTCCTAATAAAGCTGCTCATTTCTGCAGGCTATTTGAGAAAGGGGAAA 1805
QY      103 leuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuV 123
Db      1804 TCTGCTCTCAGCAGCAGACATTGCTGCACAGTACAGACACAAATGGCATGTCTATCTTG 1745
QY      123 allYsArgThrHisArgAlaIleleuPheCysLysGlnArgAspLeuLeuProGlnAsn 143
Db      1744 TGAAGAGAACACATCGGGCTAATCTGTTTGTAGCAGAGAGACTTGTACTCAAAATA 1685
QY      143 snAlaValIleuValAlaSerGlyValAlaIleSerAsnPheTyrIleArgArgAlaLeu 163
Db      1684 ATCGACTGCTGCTTGCATCTGT-GGTGTCCGAAATTAATTTAATCCGAGAGCTCTGG 1626
QY      163 IuIleuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrA 183
Db      1625 AATTTTAAAC-AAAGCAACAGTGCATTGTTGTGTCTCTCCCTCCAGATAGCACTG 1567
QY      183 sPAsnGlyIleMetIleAlaIleTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIle 203
Db      1566 ATATATGCAATTATGATGATGAGATGATATGAAAGCTACGTCGCTGGGCAATTT 1507
QY      203 euHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerL 223
Db      1506 TACATGACATAGAAAGGATCCGCTATGAAACCAAAATGTCTCTTGAGTAAACATATCAA 1447
QY      223 ysgIuValIleGlyIleAlaSerIleYsValProGlnLeuLysMetGluIle 239
Db      1446 AAGAAGTTGGAGAAAGCTTCCATTAATAAGTACCACAATTAAATAAGAGATA 1397
RESULT 12
AB075508
ID AB075508 standard; DNA; 1572 BP.
XX
AC AB075508;
XX
DT 07-NOV-2002 (first entry)
XX
DE Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.
XX
KW Murine; mouse; protease; calcium activated neutral protease type 5;
KW CAPN5; trypsinase 4; sialoglycoprotease; enzyme; genetic disease;
KW neurological; neuropsychological; psychotic illness; transgenic animal;
KW gene; ds.
XX
OS Mus musculus.
XX
PN W0200245491-A2.
XX
PD 13-JUN-2002.
XX
PP 05-DEC-2001; 2001MO-US046405.
XX

```

```

PR      06-DEC-2000; 2000US-0251803P.
PR      06-DEC-2000; 2000US-0251820P.
PR      13-DEC-2000; 2000US-0255971P.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Allen KD, Leviten MW;
XX
DR WPI: 2002-657389/70.
XX
PT Novel transgenic animal, comprising a disruption in protease target gene,
PT is useful for identifying agents that ameliorates a phenotype associated
PT with a disruption in a protease target gene.
XX
PS Example 3; Fig 7; 62bp; English.
XX
CC The present invention describes a non-human transgenic animal (I)
CC comprising a disruption in a protease target gene (PG) selected from
CC calcium activated neutral protease type 5 (CAPN5) gene, trypsinase 4 gene
CC and sialoglycoprotease-like gene. Also described is a targeting construct
CC (II), comprising a first polynucleotide sequence homologous to at least a
CC first portion of PG, a second polynucleotide sequence homologous to at
CC least a second portion of PG and a selectable marker. (II) is useful for
CC producing a transgenic mouse comprising a disruption in a protease target
CC gene, by introducing (II) into a cell, introducing the cell into a
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC useful for identifying an agent that modulates the expression or function
CC of a protease target gene, by administering an agent to (I) and
CC determining whether the expression or function of the disrupted protease
CC target gene in (I) is modulated. (I) is also useful for testing the
CC efficacy of proposed genetic and pharmacological therapies for human
CC diseases, such as neurological, neuropsychological or psychotic
CC illness. The present sequence represents murine sialoglycoprotease-like
CC gene sequence, which is used in an example from the present invention
XX
SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;
Alignment Scores:
Pred. No.: 1.28e-89 Length: 1572
Score: 870.00 Matches: 184
Percent Similarity: 82.77% Conservative: 13
Best Local Similarity: 77.31% Mismatches: 37
Query Match: 70.16% Indels: 5
DB: Gaps: 2
US-10-649-273-2_copy_176_414 (1-239) x AB075508 (1-1572)
QY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db      528 CTGTTGGKRTTAGTTCACAGAGTGTTCACATTTCCGCTCTTGGGAAAGCTTTGGACATA 587
QY      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleYsHisPProGlu 40
Db      588 GCACCGGCGACATGCTTGACAGAGTGCGAAGAAAGCTTTCTTAATCAACATCCAGAA 647
QY      41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db      648 TCTTTCAATGACAGTGGTGGAAAGCTATAGAACAGTTCGCCAAAGACGAAATGATTC 707
QY      61 HisPheAspIleYsProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db      708 CATTACTTACATCACTCACTATGACAGATCTAAGATTCGATTTTCTTTCACGGGA 767
QY      81 LeuGlnHisValThrAspLysIleIleMetLysGluLysGlnGlnGlyIleGluLys 100
Db      768 CTTCAACATATTACTGATTAAGCTAATTAACACAGAGAAAGAAAGGCAATTAAGAG 827
QY      101 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db      828 GGGCAATCTGTCTATGCTGCGACATTTGCTGCGGTACAGATGCAACAGCGTGC 887

```

QY 121 HleuValysaRgThrHisArgAlaIleuPheCysGlnArgAspLeuPro 140
CC CACCTTGGAAAGAACACATCGCGTATTCGTTTCCAGCAAAAATTGCTCT 947
Db 888
QY 141 GluAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
CC CACAGCTAACCCAGATATGATGATCTGCGAGGTGTGCAAAATGCTTACATCCGAAAA 1007
Db 948
QY 161 AlaLeuGluIleuLeuThraAsnAlaThrGlnCysThrIleuLeuCysProProArgArgLeu 180
CC GCAATGGAAATATGCGCAAAATGCAACGATGCAACGATGTTGTGCTCACT-TCAGACTG 1066
Db 1008
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
CC TGCACCTGACAAATGGGCAATATTCATGCAATGGAATGGAATTTACGTCCTGCTT 1126
Db 1067
QY 201 GlyIleLeu-HisAsp-IleGluGlyIleArgTyrGluProLysCysProLeuGlyVal- 219
CC GGCCTTTTACCATGATGATAGAGACATTCGGTATTTAAACCAATGCTCCTTTGAGTGA 1186
Db 1127
QY 220 --AspIleSerLysGluValGlyGluAlaSerIleLysValProGluLeu 235
CC 1187 GGCATTATCCCGAAGAGTTGGCAGA--ACCTTGCCCATTTAAAAAGTTA 1233
Db 1187
RESULT 13
ID AAS84622 standard; cDNA; 2734 BP.
AC AAS84622;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #20426.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG20435.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 20426; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS84197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;
XX
Alignment Scores:
Pred. No.: 4,026-58 Length: 2734
Score: 599.00 Matches: 131
Percent Similarity: 71.57% Conservative: 10
Best Local Similarity: 66.50% Mismatches: 22
Query Match: 48.31% Indels: 34
DB: 5 Gaps: 6
US-10-649-273-2_COPY_176_414 (1-239) x AAS84622 (1-2734)
QY 43 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe 62
CC ACCCTGCTAGGGGGGGCGCGCT-----AGATTCAT----- 235
Db 206
QY 63 AspIleLysProProLeuHisHisAlaLys-----AsnCys----- 74
CC -----CCTATTCTCCGATGAAAGTATGACGCTGTGTATGGAACGCTGGCAGCA 283
Db 236
QY 75 -----AspPheSerPheThrGlyLeuGluHisValThraap 86
CC TCTGCTTCTGGGAGAGCCCGAGAGCTTTTACTCATGCGGGAAGCAAAATTTGGAGATAAG 343
Db 284
QY 87 LysIleIleMetLysGlyLysGluGlu-----GlyIleGlyLysGlnIleLeu 104
CC -----GlyIleIleMetLysGlyLysGluGlu-----GlyIleGlyLysGlnIleLeu 104
Db 344
QY 105 SerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLys 124
CC TCTTCAGCAGCAGCATTTGCTGCCACAGTACAGCACACAAATGGCATGTCACTTGTGAAA 463
Db 404
QY 125 ArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGluHisAsnAla 144
CC AGAACACATCGGGCTATTCTGTTGTGTAAGCAGAGACTGTGTACTCAAAATTAATGCA 523
Db 464
QY 145 ValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIle 164
CC GACTGCTGTCATCTGCTGCTGCTGCCAGTACTTCTATATTCGACAGACTCTCGAAAT 583
Db 524
QY 165 LeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsn 184
CC TTAACAAAGCAACACAGTGCCTTTGTTGTCTCTCCCTCCAGACTATGACACTGATAAT 643
Db 584
QY 185 GlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGluHis 204
CC GGCATTATGAAATGCAATGCAATGTGATGAAAGACTACGCTGCTGGCATTTTACAT 703
Db 644
QY 205 AspIleGluGlyIleArgTyrGluProLys-----CysProLeuGlyVal 219
CC GACATTGAAAGGATCGCTATGATACCAAAATGATGTGCTCTTCCAGGGCTG 754
Db 704
RESULT 14
ID ADL86725 standard; DNA; 371 BP.
AC ADL86725;
XX
DT 20-MAY-2004 (first entry)
XX
DE DNA up-regulated in murine multipotent progenitor cells Seqid 3118.

KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
 KM HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
 KM common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
 OS Mus sp.
 XX WO2003093445-A2.
 XX PN 13-NOV-2003.
 PD 05-MAY-2003; 2003WO-US014114.
 XX PF 03-MAY-2002; 2002US-0377383P.
 XX PR (STOW-) STOWERS INST MEDICAL RES.
 XX PA L1 L;
 XX PI WPI; 2004-022656/02.
 XX DR
 XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
 PT expressed nucleic acid sequences from the discrete cell sub-populations.
 PS Claim 8; SEQ ID NO 3119; 123bp; English.
 XX
 CC This invention relates to a novel method for predicting gene potential by
 CC associating nucleic acid sequences of unknown function with particular
 CC sub-population profiles. Specifically, it refers to classifying an
 CC unknown multi-lineage affiliated gene by collecting hybridisation data to
 CC develop a gene expression map, in order to determine the discrete sub-
 CC population where it is expressed. The present invention describes methods
 CC for predicting the lineage commitment of genes associated with the self-
 CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
 CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
 CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
 CC referred to as bone marrow stem cells populations. As such, these methods
 CC can be used to identify associated multi-lineage affiliated genes and
 CC hence the underlying molecular mechanisms in physiological haematopoietic
 CC development. This polynucleotide sequence is DNA associated with a murine
 CC MPP sub population of cells of the invention.
 XX
 SQ Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
 Alignment Scores:
 Pred. No.: 2.75e-44 Length: 371
 Score: 468.00 Matches: 92
 Percent Similarity: 78.86% Conservative: 5
 Best Local Similarity: 74.80% Mismatches: 26
 Query Match: 37.74% Indels: 0
 DB: Gaps: 0
 US-10-649-273-2_COPY_176_414 (1-239) x ADL86725 (1-371)
 QY 109 AspllelalaathrrvalglnhsthrwelaCyshlsleuvallyarthrhisary 128
 Db 2 GACATTCGTCGCGGATACACAGACAGACAGGTCACCTTCGCAAAAGAAACACATGCC 61
 QY 129 AlalileuPheCyblysglnarAspleuPheupProglinaAnsnAlaValleuVala 148
 Db 62 GCTATTCTGTTTCACAGCAGAGAAATTGCTCTCTCCAGCTAACGCGATTTACTTGA 121
 QY 149 SerGlycIyValAlaSerAenPheYrllarArgAlaAleuGluIleleuThrAnla 168
 Db 122 TCTGAGAGGTGTGACAGTAACCTTGTACATCCGAAAGCATTTGGAATTCGCAATGCA 181
 QY 169 ThGlnCythrleuLeuCyepProProbrArgleuCystrAspArgGlylIleMetle 188
 Db 182 ACCGAGTCACAGTGTGTGTCTCCATCCCAAGACTGTGCACATGCAATGCAATCATATT 241
 QY 165 AlaftrpansgIylegluargleuArgAlaGlyleuGlyIleleuHIsaprllegIugly 208
 Db 242 GCATGGAATGGAATTGAANGATTACGTGCAGNCTTNGGCACTTTTACNTGANNATGAGAC 301

QY 209 lIeargIyGluProlYsCyProleuGlyValAspIleSerlysgIuValGlyGluAla 228
 Db 302 ATCCNNTATGNCACAAATNTCTNTTGGATGATTCATTTCCAGAGAGTTGCAAGAGCT 361
 QY 229 SerIleIys 231
 Db 362 GCCATPAAA 370
 RESULT 15
 ADL86726
 ID ADL86726 standard; DNA; 371 BP.
 XX
 AC ADL86726;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE DNA up-regulated in murine multipotent progenitor cells SegID 3119.
 XX
 KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
 KM HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
 KM common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
 OS Mus sp.
 XX WO2003093445-A2.
 XX PN 13-NOV-2003.
 PD 05-MAY-2003; 2003WO-US014114.
 XX PF 03-MAY-2002; 2002US-0377383P.
 XX PR (STOW-) STOWERS INST MEDICAL RES.
 XX PA L1 L;
 XX PI WPI; 2004-022656/02.
 XX DR
 XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
 PT expressed nucleic acid sequences from the discrete cell sub-populations.
 PS Claim 8; SEQ ID NO 3119; 123bp; English.
 XX
 CC This invention relates to a novel method for predicting gene potential by
 CC associating nucleic acid sequences of unknown function with particular
 CC sub-population profiles. Specifically, it refers to classifying an
 CC unknown multi-lineage affiliated gene by collecting hybridisation data to
 CC develop a gene expression map, in order to determine the discrete sub-
 CC population where it is expressed. The present invention describes methods
 CC for predicting the lineage commitment of genes associated with the self-
 CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
 CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
 CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
 CC referred to as bone marrow stem cells populations. As such, these methods
 CC can be used to identify associated multi-lineage affiliated genes and
 CC hence the underlying molecular mechanisms in physiological haematopoietic
 CC development. This polynucleotide sequence is DNA associated with a murine
 CC MPP sub population of cells of the invention.
 XX
 SQ Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
 Alignment Scores:
 Pred. No.: 2.75e-44 Length: 371
 Score: 468.00 Matches: 92
 Percent Similarity: 78.86% Conservative: 5
 Best Local Similarity: 74.80% Mismatches: 26
 Query Match: 37.74% Indels: 0
 DB: Gaps: 0
 US-10-649-273-2_COPY_176_414 (1-239) x ADL86726 (1-371)
 QY 109 AspllelalaathrrvalglnhsthrwelaCyshlsleuvallyarthrhisary 128
 Db 2 GACATTCGTCGCGGATACACAGACAGACAGGTCACCTTCGCAAAAGAAACACATGCC 61

```
Db      2  GACATTCCTGCGGGTACAGCATGCAACGCCGTCACCTTGCAGAAAAGAACACATCGC 61
Qy      129  AlalIleuPheCysIysGluArgAspIeuProGlnAsnAsnAlaValIeuValAla 148
Db      62  GCTATTCGTGTTGCAGAGCAAGAAATTGCTCTCTCCAGCTAACGCAAGTATTAGTTGTA 121
Qy      149  SerGlyIyValAlaSerAsnPheTyrIleArgArgAlaIeuGluIleIeuThrAsnAla 168
Db      122  TCTGGAGGTGTTCAGAGTACCTTGACATCCGAAAGCATTTGAAATTGTCCGCAATGCA 181
Qy      169  ThrGlnCysThrIleuIeuCysProProProArgIeuCysThrAspAsnGlyIleMetIle 188
Db      182  ACGCAGTGCACGTTGTGTGTCTCCCTCCAGAGCTGTGCACTGACCAATGGCATATGATT 241
Qy      189  AlatThrAsnGlyIleGluArgIeuArgAlaGlyIleuGlyIleIeuHisAspIleGluGly 208
Db      242  GCATGGAATGGAATTGAAGATTACGTGCNGNCTTNGGCTTTNTA.CNTGANNTAGAAGAC 301
Qy      209  IleArgTyrGluProIysCysProIeuGlyValAspIleSerIysGlyValGlyIuAla 228
Db      302  ATCCNNATATGNACCAAAATNTCNTTTTGGAGTGAACATTTCCAGAGAACTTGCAAGAGCT 361
Qy      229  SerIleIys 231
Db      362  GCCATAAAA 370
```

Search completed: February 16, 2005, 13:59:01
Job time : 468.139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:19:37 ; Search time 142.101 Seconds
(without alignments)
2752.056 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LLLAVQGVSPDLLIGKSLDI.....DISKVEGRASIKVPLQKMEI 239

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US10649273/runat_14022005_114704_16425/app_query.fasta_1.1429
-DB=Issued_Patents_NA -OPMT=fastcap -SUFFIX=rm1 -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCLIG=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10649273 @CGN_1_1_292 @runat_14022005_114704_16425 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUIDRRY -NEG SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	4	US-09-774-528-177 Sequence 177, App
2	1240	100.0	1526	4	US-10-067-443-23 Sequence 23, Appl
3	1240	100.0	2197	4	US-10-067-443-1 Sequence 1, Appl
4	1213.5	97.9	1387	4	US-10-067-443-21 Sequence 21, Appl
5	725	58.5	14364	4	US-10-067-443-20 Sequence 20, Appl
6	248.5	20.0	94750	4	US-09-596-002-38 Sequence 38, Appl
7	247	19.9	1053	4	US-09-540-236-806 Sequence 806, App
8	231	18.6	1059	4	US-09-252-991A-884 Sequence 884, App
9	231	18.6	1205	4	US-05-252-391A-801 Sequence 801, App
10	224	18.1	1830121	4	US-09-557-884-1 Sequence 1, Appl
11	224	18.1	1830121	4	US-09-643-990A-1 Sequence 1, Appl
12	221.5	17.9	996	4	US-09-902-540-6612 Sequence 6612, Ap

C	13	221.5	17.9	2582	4	US-09-902-540-503	Sequence 503, App
	14	212	17.1	1074	4	US-09-543-681A-2341	Sequence 2341, Ap
	15	205	16.5	1032	4	US-09-489-039A-2050	Sequence 2050, Ap
	16	200	16.1	1315	1	US-08-087-797-1	Sequence 1, Appl
	17	197	15.9	1092	4	US-09-107-532A-2955	Sequence 2955, Ap
	18	190	15.3	1008	3	US-08-987-121A-5	Sequence 5, Appl
	19	190	15.3	1011	3	US-08-987-121A-3	Sequence 3, Appl
	20	189	15.2	822	4	US-09-710-279-727	Sequence 727, App
	21	189	15.2	3993	4	US-09-710-279-3985	Sequence 3985, Ap
	22	187	15.1	1155	4	US-09-602-777A-147	Sequence 147, App
	23	186.5	15.0	4403765	3	US-09-103-840A-2	Sequence 2, Appl
	24	186.5	15.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl
	25	186	15.0	1006	3	US-08-961-083-51	Sequence 51, Appl
	26	186	15.0	1006	3	US-09-536-784-51	Sequence 51, Appl
	27	186	15.0	1107	3	US-09-134-001C-1072	Sequence 1072, Ap
	28	186	15.0	10974	3	US-08-961-527-214	Sequence 214, App
	29	184	14.8	1011	3	US-09-066-512-1	Sequence 1, Appl
	30	182	14.7	1011	4	US-09-583-110-216	Sequence 216, Ap
	31	182	14.7	1663	4	US-09-620-312D-6	Sequence 6, Appl
	32	181.5	14.6	432	4	US-09-328-352-261	Sequence 261, App
	33	180.5	14.6	1664976	4	US-08-916-421B-1	Sequence 1, Appl
	34	180.5	14.6	1664976	4	US-09-692-570-1	Sequence 1, Appl
	35	178	14.4	640681	4	US-09-790-988-1	Sequence 1, Appl
	36	177	14.3	1011	4	US-09-107-433-1618	Sequence 1618, Ap
	37	174.5	14.1	1230025	4	US-09-198-452A-1	Sequence 1, Appl
	38	174.5	14.1	1230230	4	US-09-438-185A-1	Sequence 1, Appl
	39	174	14.0	1026	3	US-09-149-624-1	Sequence 1, Appl
	40	173	14.0	1101	4	US-09-134-000C-1551	Sequence 1551, Ap
	41	170	13.7	15249	4	US-08-956-171E-102	Sequence 102, App
	42	170	13.7	15249	4	US-08-781-986A-102	Sequence 102, App
	43	168	13.5	3064	3	US-09-221-017B-794	Sequence 794, App
	44	159.5	12.9	42325	4	US-08-311-731A-111	Sequence 131, App
	45	145.5	11.7	3215	4	US-09-710-279-3566	Sequence 3566, Ap

ALIGNMENTS

RESULT 1
US-09-774-528-177
; Sequence 177, Application US/09774528
; Patent No. 6743619
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Yang, Yongshong
; APPLICANT: Xue, Aildong J.
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6743619e1 Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/774, 528
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
US-09-774-528-177

Alignment Scores:

Pred. No.: 1.02e-157 Length: 1416
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-09-774-528-177 (1-1416)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 586 CTGTTGGCATTAAGTTCAGAGAGTTTCAGATTTCCTTGGAAGCTTTGGACATA 645
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLysLeuIleLysHisProGlu 40
Db 646 GCACCAAGGTGACATCTTGACAGAGGTGCAAGAGACTTTTAAATAAACAATCCAGAG 705
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 706 TGCTCCACATGAGTGTGGGAAAGCCATAGAACCTTTGGCCAAACAAGAAATAGATT 765
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 766 CATTTTGACATCAAACTCCCTTGATCATGCTAAATAATTGATTTTCTTTACTGGA 825
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
Db 826 CTTCACACAGTATGATTAATAATAATGAAGAAAGAGAGAGATTTGGAAG 885
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db 886 GGGCAAAATCCGTCTTCAGCAGCAGACATGCTGCACAGTACAGCACAATGGCATGT 945
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 946 CATCTTGGAAGAAAGACATCGGGCTATTCTGTTTGTAGACAGAGACTTGTTACT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlleArgArg 160
Db 1006 CAAATTAATGACATGATCTGATGATCTGATGATGATGATGATGATGATGATGAT 1065
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 1066 GCTCTGGAATTTTAACAACGCAACAGTGCATTTGTGTCTCTCCACAGACTA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgIleLeu 200
Db 1126 TGCACTGATTAATGCGATTATGATTCATGATGATGATGATGATGATGATGATGAT 1185
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyTgIuProLysCysProLeuGlyValAsp 220
Db 1186 GGCATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1245
QY 221 HisSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 1246 AATATCAAAAGAGTGGAGAGACTTCATTAAGTACCAATTAATAAATGAGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
```

SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:

Pred. No.: 1.15e-157 Length: 1526
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-23 (1-1526)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 85 CTGTTGGCATTAAGTTCAGAGAGTTTCAGATTTCCTTGGAAGCTTTGGACATA 144
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLysLeuIleLysHisProGlu 40
Db 145 GCACCAAGGTGACATCTTGACAGAGGTGCAAGAGACTTTTAAATAAACAATCCAGAG 204
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 205 TGCTCCACATGAGTGTGGGAAAGCCATAGAACCTTTGGCCAAACAAGAAATAGATT 264
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 265 CATTTTGACATCAAACTCCCTTGATCATGCTAAATAATTGATTTTCTTTACTGGA 324
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
Db 325 CTTCACACAGTATGATTAATAATAATGAAGAAAGAGAGATTTGGAAG 384
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db 385 GGGCAAAATCCGTCTTCAGCAGCAGACATGCTGCACAGTACAGCACAATGGCATGT 444
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 445 CATCTTGGAAGAAAGACATCGGGCTATTCTGTTTGTAGACAGAGACTTGTTACT 504
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlleArgArg 160
Db 505 CAAATTAATGACATGATCTGATGATCTGATGATGATGATGATGATGATGATGATGAT 564
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 565 GCTCTGGAATTTTAACAACGCAACAGTGCATTTGTGTCTCTCCACAGACTA 624
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgIleLeu 200
Db 625 TGCACTGATTAATGCGATTATGATTCATGATGATGATGATGATGATGATGATGATGAT 684
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyTgIuProLysCysProLeuGlyValAsp 220
Db 685 GGCATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744
QY 221 HisSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 745 AATATCAAAAGAGTGGAGAGCTTCATTAAGTACCAATTAATAAATGAGAGATA 801

RESULT 3
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: SPINAL CORD, MP-1
```

```
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-067-443-1

Alignment Scores:
Pred. No.: 2,07e-157 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-1 (1-2197)

QY 1 LeuLeuAlaLeuValGlnGlyValSerApphLeuLeuGlyLysSerLeuAspIle 20
DB 756 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCCTTGGAAAGCTTTGACATA 815
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisProGlu 40
DB 816 GCACCAAGTACATGCTTGACAGCGTGCAGAAACCTTTCTTAATAAACATCCAGAG 875
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 876 TGCTCCACCATGATGCTGGGAAACCATTAACCTTTGGCCAAACAAAGAAATGATTT 935
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 936 CATTTTGCATCAAACTCCCTTGATCATGCTAAATAATGATTTTCTTTACTGGA 995
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlnGlyIleGluLys 100
DB 996 CTTCAACACGTTACTGATTAATAATAATGAAGAAAGAAAGAAAGATTTGACAG 1055
QY 101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1056 GGGCAAAATCCGCTTTCAGACAGACATCTGTCACAGTACACACACAAATGGCATGT 1115
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 1116 CATCTTGTGAAGAAACACATCGGCTATTTCTGTTTGAAGACAGAGACTTGTTACTT 1175
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspAsnPheTyrIleArgArg 160
DB 1176 CAAATAATATCAGATACGTGTCATCTGTGTGTGCGCAAGTAATCTTAATCCGAGA 1235
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1236 GCTCTGAAATTTTAAACAAAGCAACACAGTCACTTGTGTGCTCCCTCCGACTGA 1295
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1296 TGCATCTAATATGCAATTAATATGATGCAATGATTAAGAAAGCTAGCTGAGCTTG 1355
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProLysCysProLeuGlyValAsp 220
DB 1356 GGCAATTTTACATGACATAGAAAGCATCGCTATGAACCAAAATGCTCTTGAGATAGAC 1415
QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239
```

```
DB 1416 ATATCAAAAGAGTTGAGAAAGCTTCATTAAGTACCACAATTAATAAGAGATTA 1472
RESULT 4
US-10-067-443-21
; Sequence 21, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
; US-10-067-443-21

Alignment Scores:
Pred. No.: 3.87e-154 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 1
Best Local Similarity: 90.15% Mismatches: 0
Query Match: 97.86% Indels: 25
Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-21 (1-1387)

QY 1 LeuLeuAlaLeuValGlnGlyValSerApphLeuLeuGlyLysSerLeuAspIle 20
DB 549 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCCTTGGAAAGCTTTGACATA 608
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisProGlu 40
DB 609 GCACCAAGTACATGCTTGACAGCGTGCAGAAACCTTTCTTAATAAACATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 669 TGCTCCACCATGATGCTGGGAAACCATTAACCTTTGGCCAAACAAAGAAATGATTT 728
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 729 CATTTTGCATCAAACTCCCTTGATCATGCTAAATAATGATTTTCTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlnGlyIleGluLys 98
DB 788 CTTCAACACGTTACTGATTAATAATAATGAAGAAAGAAAGAAAGATTAATTTCTTA 848
QY 98 ----- 98
DB 849 ATTAGTAAGTTGAACAGATAATATTCCTGATTCCTGAATAAATAGCTGCTATTTC 908
QY 99 -----GlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
DB 909 TGCAGGTATGAAAGGGCAATTCCTGCTTTCAGACGACAGACATTCCTCCACAGTACG 968
QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB 969 CACACAAATGGCATGTCATCTTGTGAAGAAACATCGGCTATTTGTGTGAAGCAG 1028
QY 136 ArgAspLeuLeuProGlnLysAsnAlaValLeuValAlaSerGlyValAlaAspAsn 155
DB 1029 AGAGACTTGTATACCTCAAAATAATATGCAATACGTGTTGCACTGTGTGTCCAGTAAAC 1088
QY 156 PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175
```

```
|||||
Db 1089 TTCTATATCCGACAGCTCTGGAATTTTAACAAGCAACAGTGCATTTGTTGT 1148
Qy 176 ProProProAaGluCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArg 195
Db 1149 CCTCTCCCAACACTATGACATGATATATGCAATTAATGCAATGAAATGTTTGAAGAA 1208
Qy 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCys 215
Db 1209 CTACGTGTGCTGCTGGCAATTTTACATGACATGAAAGGCATCCGCTATGACCAAAATGT 1268
Qy 216 ProLeuGlyAlaAspIleSerLysGlyValGlyGluAlaSerIleLysValProGluLeu 235
Db 1269 CCTCTGGAGTAGACATATCAAAAGAAAGTGGAGAAGCTTCATAAAGTACACATTA 1328
Qy 236 LysMetGluIle 239
Db 1329 AAAATGGAGATA 1340

RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIORITY FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:
Pred. No.: 3,216-86 Length: 14364
Score: 725.00 Matches: 186
Percent Similarity: 32.86% Conservative: 0
Best Local Similarity: 32.86% Mismatches: 0
Query Match: 58.47% Indels: 380
DB: 4 Gaps: 3

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-20 (1-14364)
Qy 29 ValAlaArgAaGluSerLeuIleLysHisProGluCysSerThrMetSerLysGlyLys 48
Db 11840 GTGCGAAGAAAGACTTTCTTTAATAAACAATCCAGAGTCTCCACATAGATGTGGGAAA 11899
Qy 49 AlaIleGluHisLeuAlaLysGluGlyAsnArgPheHisAspIleLysProProLeu 68
Db 11900 GCCATAGAACATTTGGCCAAACAAGAAATAGATTTCATTTGACATCAAACTCCCTTG 11959
Qy 69 HisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGluHisValThrAspLysIle 88
Db 11960 CATCATGCTAAATAATGATTTTCTTTTACGTGACCTTCACACAGTTACTGATAAATA 12019
Qy 89 IleMetLysLysGlyLysGlyLys----- 96
Db 12020 ATATATGAAAAAGAAAAAGAGAGATATTTCTTAATTAGTAAAGTTCACACAGTAAT 12079
Qy 97 -----GlyIleGlyLysGlyLysIle 103
Db 12080 ATTCCTGATTTGTGCTAAATAATAGCTGCTATTTCTGCAAGGTATTTGAGAGGGGCAAA 12139
Qy 103 eleuSerSerAlaAlaAspIleAlaIleAlaThrValGluHisIleMetAlaCysHisLeuVal 123
```

```
|||||
Db 12140 CCTGCTTCAGACAGACAGATGCTGCGACAGTAGACGACACAAATGGCATGTCTTGT 12199
Qy 123 LysArgThrHisArgAlaIleLeuPheCysLysGlyAsnArgProGluHisPhe 143
Db 12200 GAAAGAAACACATCGGGCTATTCGTTTGTGAAGCAGAGACTGTGTAACCTTAATAA 12259
Qy 143 nAlaValLeu----- 146
Db 12260 TGCAGTACTGTAAGTTTATCTCATTTTATAGTAATAGTACACTTGCATATGTAC 12319
Qy 146 ----- 146
Db 12320 TTTTTCCTCAAGACCTTGACCTTGTTGTAGATGAACAGATCTTATGCTTATGCTAG 12379
Qy 146 ----- 146
Db 12380 CCTGACAGTATGAATATATGACAGATGAGAAAGACTAAACGCCATTTCTGTACTAGTT 12439
Qy 146 ----- 146
Db 12440 TGGTACCTTTATGAGACAGCTGTATAGCTTCTATGACACATAGCTAATTTTCATCTT 12499
Qy 146 ----- 146
Db 12500 CTGTTGATTTAAAAAGAGGGCTTACATATAAAGAAAGTAATAGCAGTAACGTGCTACCT 12559
Qy 146 ----- 146
Db 12560 ATTTTATGAAAAATAGTNGAATTTCTTCATCTTTGATGAATAATCCCTTGTGTTGTGT 12619
Qy 146 ----- 146
Db 12620 TTTTATTAAGCCAGTCAAAATTTAGCAGTGGAGGTGTATTCAACTTTCGTGACACT 12679
Qy 146 ----- 146
Db 12680 AATGTTGATTAAGTTCGTATATCCATATTTGTACAGCCAAATCCCTTAATTGTG 12739
Qy 146 ----- 146
Db 12740 CTTAAAAACCTTGACAAACATCCTGTTTAACTGATCTTAACTTTATCATTTAAAAAT 12799
Qy 146 ----- 146
Db 12800 TATTAACATAAGTGGAAAAATGTTAAATGTATGATTAATCATAGATGAATTTTACATGG 12859
Qy 146 ----- 146
Db 12860 ATATCAAGATTAATTTTTCAGAGTTATGTAGTAATAATGCACAAATAATAAATTTTC 12919
Qy 146 ----- 146
Db 12920 AGGGTCTAAATAGTGTACTATGATGAATTAATTAATAATATTATGATGAAGGT 12979
Qy 146 ----- 146
Db 12980 TGGAGAAAAATATACAAAAATGCTAGTAATGTTGTATGCTATTAGAAATTAATTAAT 13039
Qy 146 ----- 146
Db 13040 TTTTCTTCCAAATTTTATTAATACATAGATATGCTACTGCCATTAACCATCTCAAAAT 13099
Qy 146 ----- 146
Db 13100 GGGATAGTTTATTTGTTAATGCTATATTTTTCACAGTTTATAGACGTTGT 13159
Qy 147 -----ValAlaSerGlyIle 151
Db 13160 TCAATACATATATGATAGTATTTTGTGTTTCTCAATTCCTTCAGGTGCACTGTGTG 13219
Qy 151 yValAlaSerAsnPheTyrIleArgArgAlaLeuGlyIleLeuThrAspAlaThrGlnCys 171
```

Db	13220	TCGCGAAGTAACCTTCTGATATACGACAGCCTGGAATTTTAAACAAGCAACAGTG	13279
Qy	171	ethleuleuCySPProProPtoAigleuCyethAspansglylleme1lealATP--	190
Db	13280	CACCTTGTGTGTGCTTCCTCCGACACTATGACATGATATATGCAATTATGATGCAATGTA	13339
Qy	190	-----	190
Db	13340	AGCCACAGATATATACGTCTTCACTCACTAACTATGTAAATATTAATGCCATTTTATCAT	13399
Qy	190	-----	190
Db	13400	ACTAAGCCTTCTTCTTCAGATCTTGAGGCTATGATTTTATTTAATGCTTCTATTTA	13459
Qy	191	--AsnglyllegluaagleuAaglaaglyleuglylleuHieaspillegluglyllea	210
Db	13460	GGAATGGTATTTGTAAGACATACGTGCTGCTTGCGCATTTTACATGACATGAGGCAATCC	13519
Qy	210	IGTYrGluProlys 214	
Db	13520	GCTATGACCAAAAG 13533	
RESULT 6			
US-09-596-002-38			
Sequence 38, Application US/09596002			
Patent No. 6632636			
GENERAL INFORMATION:			
APPLICANT: Lagace, Robert, E.			
APPLICANT: Paterson, Chandra			
APPLICANT: Berg, Kim, L.			
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME			
FILE REFERENCE: PM-0008-4 US			
CURRENT APPLICATION NUMBER: US/09/596,002			
CURRENT FILING DATE: 2000-06-16			
PRIOR APPLICATION NUMBER: 60/140,121			
PRIOR FILING DATE: 1999-06-18			
NUMBER OF SEQ ID NOS: 41			
SOFTWARE: PERL Program			
SEQ ID NO 38			
LENGTH: 94750.			
TYPE: DNA			
ORGANISM: M. catarrhalis			
FEATURE:			
NAME/KEY: misc_feature			
OTHER INFORMATION: incyte template ID No. 6632636 38			
PUBLICATION INFORMATION:			
US-09-596-002-38			
Alignment Scores:			
Pred. No.: 3.02e-20			
Score: 248.50			
Percent Similarity: 48.67%			
Best Local Similarity: 32.30%			
Query Match: 20.04%			
DB:	4	Gaps:	7
US-10-649-273-2_COPY_176_414 (1-239) x US-09-596-002-38 (1-94750)			
Qy	1	LeuleuAlaleuValGInglyValserAspPheleuleuenglylyserSetleuaspille	20
Db	24305	ATGCTGCGTCGCGCCAGATGCTGCGCGCTGATACATATATGGCGAGCTTATGCATAT	24364
Qy	21	AlAProGlyAspMetleuaspLyValAlaArgArgleuSerleuileuHieProGlu	40
Db	24365	GCGGCGGGGGAATGCTTTATATAAAGCGCAAAATGCTCAACTG---CCCTATCT---	24418
Qy	41	CySerThmetSerGlyglylybalaileGluHieleuAlalyGInglyAsnArgPhe	60
Db	24419	-----GCTGCGCCAAATATGAAAAAATTAAGCAAAAAAGGCAACCCACAC	24463
Qy	61	HispheAspilleysProProleuHieHialalyAsnCySaAPheSerPheThrcly	80
Db	24464	GCTATGACCTGCGCAAGACCCACATGCAAGCAT---AAAGCGCTGATTTTTCGTTAGTGCC	24520

QY 81 LeuGlnHisValThrAspIleLeuMetLysGluLysGluGly1LeuLys 100
Db 24521 ATGAAAACCGCCATTATCATATCTCATCAAGACAACAAAGCCCAAAGGACCCC--- 24577
QY 101 GlyIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHISThrMetAlaCys 120
Db 24578 -----GCCAAGACGACGACATCGCCGCAAGCTTTGGTAGTATGCGGTGGAT 24625
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 24626 ACTTGTCATAAAAAAATGACCAAAACATACAGATGCAGGCATTCGCCAG----- 24676
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
Db 24677 -----CTGGTCGTCGACGGGGGGCTCTCTCCCAATCAGATGCTACGCCG 24721
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 24722 ACCCTGACCCGACGCTCCGCAATCGATGCGTCGGTAGTCTATGCCCCGACGACTA 24781
QY 181 CysThrAsnArgGlyLysMetIleAlaTrpAsnGlylIleGluArgLeuAlaGlyLeu 200
Db 24782 TGCACGGAATAAATGTCGATGATGCGCTATGCTGCTTTGTCGGCTCACGCTGACAG 24841
QY 201 GlyIleLeuHisAspIleGluGlylleArgTyrlleProLysCysProLeuGlyValAsp 220
Db 24842 TCG-----GAGCAGCTTGGC-GTTTCGCTGTATTC-----CCGATGGAGTATAGC 24885
QY 221 IleSerLysGluValGly 226
Db 24886 GACGCTTGCGCTATCGGC 24903

RESULT 7
US-09-540-236-806
; Sequence 806, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SBO ID NO 806
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-806

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:
3,22e-23	247.00	46.85%	31.53%
Length: 1053	Matches: 70	Mismatches: 34	Indels: 38
Gaps: 6			

Query Match: 4 19.92% DB: 6

US-10-649-273-2_COPY_176_414 (1-239) x US-09-540-236-806 (1-1053)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGluGlyLysSerLeuAspIle 20
Db 445 ATGCTGCGTGGCGGCGGATGCTGCGCGGTATCAGATATTTGGCGAGTCCATCGATGAT 504
QY 21 AlaProGlyAspMetLeuAspLysValAlaIleArgArgLeuSerLeuIleLysHisProGlu 40
Db 505 GCGGAGGCGGATGCTTGTATAAACGGCAAAATCTCAAACTG---CCCTATCTCT--- 558
QY 41 CysSerThrMetSerGlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 559 -----GATGCTCCAATATCGAAAAAATTAGCCAAAAACGGCAACCCACAC 603

Oy 61 HispRheAspRLeuYsPProFoleuHnHhAlaYsAsnCyAspRPhseRheTngly 80
 Db 604 GCGTATGAGCTGCGCAGACCGATGACAGAT--AAAGGAGTGAATTTTTCGTCAAGTGC 660
 Oy 81 LeuGlnHleValHthrAspYsleIleTemeYsYsGluYsGluGluYleGluYs 100
 Db 661 ATGAAACCGCGCATTCATTAATCTCATCAAAACACGCCAAGCCCAAGCGACCT--- 717
 Oy 101 GlyGlnIleuSerSerAlaAlaAspRleAlaAlaThrValGlnHleThrMetAlaCys 120
 Db 718 -----GCCACACGACGACACATCCCGCAACCTTGAAGTATCGGTGTGGAT 765
 Oy 121 HisLeuValYAspRgThrHsAspAlaAlaLeuPheCysYsGlnHthrAspLeuPro 140
 Db 766 ACTTGTGCAAAAATGACACCAAGCCTACAGATGACAGGACGATTCGCCAG----- 816
 Oy 141 GlnAsnAspAlaValLeuValAlaSerGlyGlyValAlaSerAspHetyrIleArgTg 160
 Db 817 -----CTGCTGTGCGCAGGGGGCGCTCTGCGCAACAGACGCTACGCCGC 861
 Oy 161 AlaLeuGluIleLeuThrAspAlaThrGlnCysYsThrLeuLeuCysPProFolArgLeu 180
 Db 862 ACCCTGACCGAGACGCTGCGCGCAATGATGATCGTGGGTACTAGTACGCCGACCGAGCTA 921
 Oy 181 CysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluArgLeuArgAlaGly--- 199
 Db 922 TGACACGATTAATGTGTGCGATGATGCTTACGCTGTCTTTGTGGCTAAGCCGTGGGCGAG 981
 Oy 200 -----LeuGlyIleLeu 203
 Db 982 TCGGATGACTTGGCGGCTTCCGTATTCCTCCGATGGAGATGACATGACTTGGTATCGAA 1041
 Oy 204 HisAsp 205
 Db 1042 TATGAT 1047
 RESULT 8
 US-09-252-991A-884/c
 ; Sequence 884, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 884
 ; LENGTH: 1059
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-884
 Alignment Scores:
 Pred. No.: 4.81e-21 Length: 1059
 Score: 231.00 Matches: 73
 Percent Similarity: 49.54% Conservative: 34
 Best Local Similarity: 33.80% Mismatches: 87
 Query Match: 18.63% Indels: 22
 DB: 4 Gaps: 8
 US-10-649-273-2_COPY_176_414 (1-239) x US-09-252-991A-884 (1-1059)
 Oy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuGlyYsSerLeuAspIleAla 21
 Db 621 TTGGTGGCGGGGTGACCGGTATGCGCGCTACCAAGTTCCTTGGCGAATCGGTGACATGCC 562
 Oy 22 ProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleYsHisProGluCys 41

[illegible]

```

D6 604 TTGGTGGGGGAGACGGATACGGCGCGCTACAGATTGCTTGGGAAATCGGTGACATGCG 663
OY 22 ProgliaPmMeLleuArbLyValAlaArgArgLeuSerLeuIlleYnHlProGluCys 41
D6 664 GCGCGCGAAGCGCTTCGACAGACCGCCAGCGATGATCGGCTTG---GGCTATAC----- 714
OY 42 SerThmetsSerGlyValYnAlaIlleGluHlsluAlaYnGlnGlyYnAsnArgHnHs 61
D6 715 -----GGTGTTCGGAAATTCGCGCGCGCTCGCGGACCGCGACCTCTTGCGCGC 762
OY 62 PheArpIlleYnProProLeuHlshlValYnAsnCysAspPheSerPheThrlYnLeu 81
D6 763 TTCGTGTTCCCGCGCGCGATGACGATGCGCGCGCGCTGAACTTCAAGCTTCAAGCGGCTC 822
OY 82 GlnHlslValThrArpLyIlleIlleMetYnYnGluYnGluGluGlyIlleGluYnGly 101
D6 823 AAGACCTTACCTTAC---ACCTGGACAGCGTTCGATGAGCGCGGACGACGACGAG 879
OY 102 GlnIlleuSerSerAlaAlaArpIlleAlaIlleValGlnHlshThmAlaCysHnS 121
D6 880 CAG-----ACCGCGTGCACATCGCGCTTCGAGCGCTTCAACACGCGGTGTGAGAC 930
OY 122 LeuValYnArgThrlshArgHlAlleuPheCysYnGlnArgAspPheLeuProGln 141
D6 931 CTGCGTATCAAGTGCCTCGCGCGCTTG-----AAGCAACCGGCGCTG-----AAG 975
OY 142 AsnAsnAlaValLeuValAlaSerGlyValAlaAsnAspPheThrlleArgHlAla 161
D6 976 AAC-----CTGGTATGTCGCCGCGGTGTCAACGCGCAACGAGCGCTCGCAGCGAG 1022
OY 162 LeuGluIlleuThrAsnAlaThrGlnCysThrlleuLeuCysProProArgLeuCys 181
D6 1027 CTGGAAGAAGATGCTCGCGGAAATGAAAGGGGACAGGTTCCTACGCGCGCGCTTCTGC 1086
OY 182 ThrAspAsnGlyIlleMetIlleAlaThrAsnGlyIlleGluArgLeuArgHlAlaGly 201
D6 1087 ACCGACATGCGCGCATATCGCTTACCGCGCGCTGCGACGCGCTGCTGCGCGCGC----- 1144
OY 202 IlleuHlshArpIlleGluGlyIlleArgTyrluProLysCysProLeu 217
D6 1141 ---CAGCATGACGGCGCGCATGACGCTTCAGCGCGCGCGCTGCGCATG 1185

RESULT 10
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557, 884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 06/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks

```

```

1      REGISTRATION NUMBER: 41,971
2      REFERENCE/DOCKET NUMBER: PB186P3
3      TELECOMMUNICATION INFORMATION:
4      TELEPHONE: 301-309-8504
5      TELEFAX: 301-309-8439
6      INFORMATION FOR SEQ ID NO: 1:
7      SEQUENCE CHARACTERISTICS:
8          LENGTH: 1830121 base pairs
9          TYPE: nucleic acid
10         STRANDEDNESS: double
11         TOPOLOGY: linear
12
13     SEQUENCE DESCRIPTION: SEQ ID NO: 1:
14
15     US-09-557-884-1
16
17     Alignment Scores:
18     Pred. No.:          7.8e-15          Length:          1830121
19     Score:              224.00           Matches:           67
20     Percent Similarity: 47.50%           Conservative:    28
21     Best Local Similarity: 33.50%        Mismatches:      83
22     Query Match:        18.06%          Indels:         22
23     DB:                  4              Gaps:           6
24
25     US-10-649-273-2_COPY_176_414 (1-239) x US-09-557-884-1 (1-1830121)
26
27     Oy      2      LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
28     Db      552791  TTATGTCGCTGATGATGTTAGAAAATTTGAAGTGAAGAAATCTATATGATGATGCT 5527322
29     Oy      22      ProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleLysHisProGlyCys 41
30     Db      552731  GCTGCGAAGCCTTGATTAACACGAAATTAATCTTGACTA---GATATATCA----- 552681
31     Oy      42      SerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61
32     Db      552680  -----GGTGGCGCGGCACCTTCTCGTTTACGGAAAAAGTACGCCAAATTCGT 5526333
33     Oy      62      PheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeu 81
34     Db      552632  TTGCATATTCACGTCACAAATGACAGATCGTCAGGCGTTGAATTTTATGTTTCTGCGTTTA 5525737
35     Oy      82      GlnHisValThrAspLysIleIleMetLysGlyLysGluGly-----IleGlu 99
36     Db      552572  AAAACATTGGCCGCAATACGATTATCAAGCAATTTAAACAGAGGCGCAACTGATAGAC 5525133
37     Oy      100     LysGlyGlnIleLeuSerSerLysAlaAspIleAlaIleAlaThrValGlnHisThrMetAla 119
38     Db      552512  CAA-----ACTAAACACAGATATTCCTTATGCTTCCAGATGCGGCGTGC 552468
39     Oy      120     CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgPheLeu 139
40     Db      552467  GATACTCTGGCC-----ATTAATATTAAG---CGTGCAATGAAA 5524323
41     Oy      140     ProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArg 159
42     Db      552431  GAAACAGCGCTATAACGTTTATGATTTGGCGAGGGGTAGCGCAATTAACAACTCCGA 5523722
43     Oy      160     ArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPheArg 179
44     Db      552371  GAAACGCTGGCGCACTTAATGCAAAATTTAGGTGGCGAAGTGTTTATCTCAACCTCAA 5523121
45     Oy      180     LeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGly 199
46     Db      552311  TTTTGTCACAGATTAAGTGGCGATGATGCTTACACAGGTTTTTTTACGTTTAAACAAAGG 5522522
47
48     RESULT 11
49     US-09-643-990A-1/C
50     ; Sequence 1, Application US/096433990A
51     ; Patent No. 6528289
52     ; GENERAL INFORMATION:
53     ; APPLICANT: Robert D. Fleischmann
54     ; Mark D. Adams
55     ; Owen White
56     ; Hamilton O. Smith
57

```

J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Alignment Scores:
Pred. No.: 7.8e-15 Length: 1830121
Score: 224.00 Matches: 67
Percent Similarity: 47.50% Conservative: 28
Best Local Similarity: 33.50% Mismatches: 83
Query Match: 18.06% Indels: 22
Gaps: 4
DB: 6
US-10-649-273-2_COPY_176_414 (1-239) x US-09-643-990A-1 (1-1830121)
QY 2 LeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLySerLeuAspIleAla 21
DB 552791 TTAAGTGGGTGCGATGCGTGAAGAAATATGAAGATGAGATGATATGATGT 552732
QY 22 ProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleLyHisProGlyLucy 41
DB 552731 GCTGGCGAAGCCTTGTGATTAACAGCAAAATTAAGTGAAGTGA---GATTATCCA----- 552681
QY 42 SerThrMetSerGlyGlyLyValAlaIleGluHisLeuAlaLySerGInGlyAspArgPheHis 61
DB 552680 -----GGTGGCGCGGCACTTCTCGTTTGAAGGAAAGGTAAGCGCAAAATCGT 552633
QY 62 PheAspIleLeuValProPheLeuHisIleAlaLyAsnCyAspPheSerPheThrGlyLeu 81
DB 552632 TTTCACATTTCCACGCGCAATACAGATGTCGAGCCCTTGATTTTATGTTTCTGCTTTA 552573
QY 82 GlnHisValThrAspIleIleIleMetLySerGlyGlyGluGlyLy-----IleGlu 99
DB 552572 AAACACATTTGCGCAATATACAGTATATCAAGCAATTAATAAGAGGCGAAGCTGATAGG 552513
QY 100 LySerGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAla 119

DB 552512 CAA-----ACTTAACAGATATATGCTTATGCTTCCAAAGATCGGTGGTG 552468
QY 120 CysHisLeuValLyArgThrHisArgAlaIleLeuPheCysLyGlnArgAspLeuLeu 139
DB 552467 GATACCTTGGCC-----ATTAATGTAG---CGTGCATTGAAA 552432
QY 140 ProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheThrIleArg 159
DB 552431 GAAACAGGCTATTAACAGTTTATGATGCGGAGAGGGGTAGCGCAATATAAAACTCCCA 552372
QY 160 ArgAlaLeuGluIleLeuThrAsnAlaThrGlnCySerIleLeuCyAspProProArg 179
DB 552371 GAAACGCTTGGCGCACTTATATGCAAAATTAAGTGGGAGAGGTATTATCTCAACTCA 552312
QY 180 LeuCySerThrAspAsnGlyIleMetIleAlaIleArgGlyIleGluArgLeuArgAlaGly 199
DB 552311 TTTTGTACAGATTAATGTCGATGATGCTTACACAGGTTTATTAAGTTTAAACAAAGGT 552252
RESULT 12
US-09-902-540-6612
/ Sequence 6612, Application US/09902540
/ Patent No. 6833447
/ GENERAL INFORMATION:
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Wiegand, Roger C.
/ TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
/ FILE REFERENCE: 38-10(15849)B
/ CURRENT APPLICATION NUMBER: US/09/902,540
/ PRIOR FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: 60/217,883
/ PRIOR FILING DATE: 2000-07-10
/ NUMBER OF SEQ ID NOS: 16825
/ SEQ ID NO 6612
/ LENGTH: 996
/ TYPE: DNA
/ ORGANISM: Myxococcus xanthus
US-09-902-540-6612
Alignment Scores:
Pred. No.: 8.47e-20 Length: 996
Score: 221.50 Matches: 70
Percent Similarity: 47.50% Conservative: 25
Best Local Similarity: 35.00% Mismatches: 74
Query Match: 17.86% Indels: 31
Gaps: 4
DB: 8
US-10-649-273-2_COPY_176_414 (1-239) x US-09-902-540-6612 (1-996)
QY 5 ValGInGlyValSerAspPheLeuLeuGlyLySerLeuAspIleAlaProGlyAsp 24
DB 433 GTGACAGCGCTACGGGACAGTACGGGCTGTGGGACAGACGCGGCGCGCGAG 492
QY 25 MetLeuAspIleValAlaArgArgLeuSerLeuIleLyHisProGlyLucySerThrMet 44
DB 493 GCATATGACAAACCGCTGATCTCGGCGCTG---CCGATATCG----- 534
QY 45 SerGlyLyValAlaIleGluHisLeuAlaLySerGInGlyAsn-----ArgPhe 60
DB 535 ---GGTGGCGCGGCACTTCTCGTTTGAAGGAAAGGTAAGCGCAAAATCGT 552633
QY 61 HisPheAspIleLeuValProPheLeuHisIleAlaLyAsnCyAspPheSerPheThrGly 80
DB 592 -----CCGCGCGCGGCTGCGGCGCAACTTGTGAAGTGTCTTCTCGG 636
QY 81 LeuGlnHisValThrAspIleIleIleMetLySerGlyGlyGluGlyLyIleGluLy 100
DB 637 TTGAAG-----ACGCGGTGCTGCACCAAGTGCAGAAAGCAAGCGGTGCGCGAG 684
QY 101 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAla 120


```

Db      685 GGGCAGCGCGT-----GCGGATTTCGCGCTCTTCCAGAGAGCGCTGCGGAC 735
Qy      121 HisleuVallybArGThhIbArGAlaileuPheCylybGlnArgbLeuPro 140
Db      736 GTGCTGTGGAAGAG-----CTGCTGCGCGCGCGCGCGCGCTTG----- 774
Qy      141 GlnbAmnAlaValleuValAlaSerGlyValAlaSerbAmPheTyrlleaArg 160
Db      775 ---GGCCACAAGCAGTGTGCTGTGCGCGCGCTGCGCGCACTCGCGCGTGGCGGCA 831
Qy      161 AlaLeuGlnilleuThraAmAlaThrGlnCysThrlleuLeuCybProProArgleu 180
Db      832 CTGTGTAGCGCGCAGCGCAGAGCGGCGGTGAACATGTTCTGCCCCGCTGCGCGCTG 891
Qy      181 CyThraSPaenglylleMetilleaIATrPaenglylleGluArgleuArglaGlyleu 200
Db      892 TGCAAGACATGCGCGCATGATTCGCGTGGCGGGATATAGCGGTACCGCGCGCGCTG 951

RESULT 13
US-09-502-540-503/c
; Sequence 503, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 503
; LENGTH: 2582
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-502-540-503

Alignment Scores:
Pred. No.: 3,98e-19 Length: 2582
Score: 221.50 Matches: 70
Percent Similarity: 47.50% Conservative: 25
Best Local Similarity: 35.00% Mismatches: 74
Query Match: 17.86% Indels: 31
DB: 4 Gaps: 8

US-10-649-273-2_COPY_176_414 (1-239) x US-09-502-540-503 (1-2582)
Qy      5 ValGlnGlyValSerbPheLeuLeuLeuGlylybSerLeuAPrilleAlaProGlybAP 24
Db      1096 GTGCAAGCGCTTACGGGAGTACCGGCTGTGGGACAGCGCGCAGCAGCGCGCGCGAG 1037
Qy      25 MetLeuAPrlybValAlaArgArgLeuSerleuLeuLybshIProGluCybSerThmet 44
Db      1036 GCATATACAAAGCCGCTGCGATCTCGAGCTG---CCGATATCG----- 995
Qy      45 SerGlylybValAlaIleGlnIshleuAlaLybGlnGlyAen-----ArgPhe 60
Db      994 ---GGTGGCGCGCCATGACAGTGGCGCGACAGCGGAGGCCGCGAGCCATCGCGCTTC 938
Qy      61 HisPheAPrlybProProLeuIshIshAlaLybAmCybAPrPheSerPheThrGly 80
Db      937 ---CCGCGCGCGCTGCGGCGCACAATTCTGACGTGCTTCTTCGCGG 893
Qy      81 LeuGlnIshValThraAPrlybIleleuLybGlybGluGluGlylleGlylyb 100
Db      892 TTGAG-----ACGCGGTGTGCGACCACTGCGACAGCAGCGCGTGGCGGAC 845
Qy      101 GlnIleuSerSerAlaAlaAPrilleAlaIshAlaThraValGlnIshThmetAlaCys 120
Db      844 GGGAGGCGCGT-----GCGGATTTCGCGGCTCTTCCAGAGAGCGCGTGGCGGAC 794

```

```

Qy      121 HisleuVallybArGThhIbArGAlaileuPheCylybGlnArgbLeuPro 140
Db      793 GTGCTGTGGAAGAG-----CTGCTGCGCGCGCGCGCGCGCTTG----- 755
Qy      141 GlnbAmnAlaValleuValAlaSerGlyValAlaSerbAmPheTyrlleaArg 160
Db      754 ---GGCCACAAGCAGTGTGCTGTGCGCGCGCTGCGCGCACTCGCGCGTGGCGGCA 698
Qy      161 AlaLeuGlnilleuThraAmAlaThrGlnCysThrlleuLeuCybProProArgleu 180
Db      697 CTGTGTAGCGCGCAGCGCAGAGCGGCGGTGAACATGTTCTGCCCCGCTGCGCGCTG 638
Qy      181 CyThraSPaenglylleMetilleaIATrPaenglylleGluArgleuArglaGlyleu 200
Db      637 TGCAAGACATGCGCGCATGATTCGCGTGGCGGGATATAGCGGTACCGCGCGCGCTG 578

RESULT 14
US-09-543-681A-2341
; Sequence 2341, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2341
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2341

Alignment Scores:
Pred. No.: 1.86e-18 Length: 1074
Score: 212.00 Matches: 66
Percent Similarity: 44.95% Conservative: 32
Best Local Similarity: 30.28% Mismatches: 90
Query Match: 17.10% Indels: 30
DB: 4 Gaps: 7

US-10-649-273-2_COPY_176_414 (1-239) x US-09-543-681A-2341 (1-1074)
Qy      2 LeuAlaLeuValGlnGlyValSerbPheLeuLeuGlylybSerLeuAPrilleAla 21
Db      475 TTAATTAGTGTAAACAGGATTGGGAAATATACCTGTTAGTGAGTCTATGATGATGCT 534
Qy      22 ProGlybAPrMetLeuAPrlybValAlaArgArgLeuSerleuLeuLybshIProGluCyb 41
Db      535 GCGGTGAAGCATTTGATTAACAGCCAGCATATTCGGGCTT---GATTATATCC----- 585
Qy      42 SerThmetSerGlylybValAlaIleGlnIshleuAlaLybGlnGlylybshArPheIsh 61
Db      586 ---GGCGCGCGCTGTTTATCAAAAATGCGACACACAGGTGTAGAGAGCGT 633
Qy      62 PheAPrillebProProLeuIshIshAlaLybAmCybAPrPheSerPheThrGlyleu 81
Db      634 TTGTGTTTCTCTGTCCTCAAGACAGACAGACCGCGCATCTTAAGTTAGTTCTGCTGTTA 693
Qy      82 GlnIshValThraAPrlybIleleuLybGlybGluGluGlylleGlylyb 101
Db      694 AAAACCTTTCGCGCTAATCTATTCGTCAAAAGATTCAGAG----- 738
Qy      102 GlnIleuSerSerAlaAlaAPrilleAlaIshAlaThraValGlnIshThmetAlaCys 121
Db      739 ---CAAACTGACAGATATTCGCGGCTTTTGAAGATCCGATGATGATCT 789
Qy      122 LeuVallybArGThhIbArGAlaileuPheCylybGlnArgbLeuProGln 141

```

```
Db 790 TTGGCAATAAATGTCGACGA-----TTAGAGCAA 822
Qy 142 AsnAen---AlaValLeuValAlaSerGlyValAlaSerAsnPhetYrIleAArg 160
Db 823 ACAGCGTTTAAACGCTTGAATGCTGCGGCGGATGCTAACCGTACCTTACCGCC 882
Qy 161 AlaleuGluIleLeuThraSnaIaThrGlnCystrIleuLeuCySPProPProAArgLeu 180
Db 883 AAAATGCGGATGATATGGAACAACCTCGAGCGGAAGTCTTTATGCTGCGCTGAGTTA 942
Qy 181 CystrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 943 TGTACCGATATGCGGATGCTTGGCGGCGATGATCCGTTTAAAGTGCTAC 1002
Qy 201 ---GlyIleLeuHisAspIleGluGlyIleArgTyrgIuProLysCySPProleu 217
Db 1003 GAGGCGGCTTTA-----GCGGAGACATGAGACACGTTGCGCTTTA 1044

RESULT 15
US-09-489-039A-2050
; Sequence 2050, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2050
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2050

Alignment Scores:
Pred. No.: 1.55e-17 Length: 1032
Score: 205.00 Matches: 66
Percent Similarity: 43.58% Conservative: 29
Best Local Similarity: 30.28% Mismatches: 93
Query Match: 16.53% Indels: 30
DB: 4 Gaps: 7

US-10-649-273-2_COPY_176_414 (1-239) x US-09-489-039A-2050 (1-1032)
Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
Db 442 TTGATTACCGTCAACCGGATTTGGTCAATGAACTGCGGCGAGATGCAATGACATGCG 501
Qy 22 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 41
Db 502 GCGGCGCAAGCCTTGAATGACGCGCAAGCTGCGGACTG---GATTATCC----- 552
Qy 42 SerThreSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61
Db 553 -----GCGGCGCGGATGCTGCAAAATGCGCTGCCAGGCGCACGAGCGCGC 600
Qy 62 PheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeu 81
Db 601 TTTGTCTTCCCGCGCGCATGACCGACCGCTCGGGGCTGAGCTTCACTTCTCGGCGCTG 660
Qy 82 GlnHisValThrAspLysIleIleMetLysGluLysGluGluGlyIleGluLysGly 101
Db 661 AAGACCTTGGCGCGCAACACCATTCGACGACGCAACGCGGACGATGAC----- 705
Qy 102 GlnIleuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHis 121
Db 706 -----CAACCGCGCGGACATGCGCGGCGGCTTTGACGATGCGGCTCGATACG 756
Qy 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141
```

```
Db 757 CTGATGATTAATATGTCGCGCGCG-----CTGAGACAA 789
Qy 142 AsnAen---AlaValLeuValAlaSerGlyValAlaSerAsnPhetYrIleAArg 160
Db 790 ACCGCGTTTAAACGCTTGAATGCTGCGGCGGATGCTAACCGTACCTTACCGCC 849
Qy 161 AlaleuGluIleLeuThraSnaIaThrGlnCystrIleuLeuCySPProPProAArgLeu 180
Db 850 AAGCTGCGGATGATGCAAAAACGCGCGGAGGTGTTTACGCGCGCTGAGTTT 909
Qy 181 CystrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 910 TGTACTGACACGCGCGGATGATCGCTACGCGCGATGATGCTGCTGCAACCGCGCGC 969
Qy 201 GlyIleLeuHisAspIleGlu---GlyIleArgTyrgIuProLysCySPProleu 217
Db 970 -----AAAGCCGAGCTCGGCGTACGCTTGGCGCGCGCTGCGCGCTG 1011
```

Search completed: February 16, 2005, 22:38:34
Job time : 1516.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 18:03:13 ; Search time 478 Seconds

(without alignments)
2950.230 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414

Perfect score: 1240
1 L1ALVQGVSPFLIGKSLDI.....DISKEVGEASIKVPLQKMEI 239

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Delop 6.0 , Delext 7.0	

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+g2n.model -DBV=x1h
-Q=/cgn2_1/USPTO.epool/US10649273/runat_14022005_114706_16497/app_query.fasta_1.1429
-DB=Published Applications NA -QFMT=fastap -SUFFIX=nmph -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649273@CN_1_1_1053@runat_14022005_114706_16497
-NCPU=6 -ICPU=3 -NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	1240	100.0	1416	17	US-10-120-988-177
2	1240	100.0	1526	14	US-10-067-443-23
3	1240	100.0	1526	17	US-10-649-273-23
4	1240	100.0	1526	17	US-10-649-273-23
5	1240	100.0	2197	14	US-10-067-443-1
6	1240	100.0	2197	17	US-10-649-273-1
7	1240	100.0	2197	17	US-10-651-722-1
8	1213.5	97.9	1387	14	US-10-067-443-21
9	1213.5	97.9	1387	17	US-10-649-273-21
10	1213.5	97.9	1387	17	US-10-651-722-21
11	1203	97.0	1245	14	US-10-012-140-6
12	1203	97.0	1820	14	US-10-012-140-4
13	1059	85.4	2208	17	US-10-094-749-400
14	1059	85.4	2890	18	US-10-723-860-7447
15	725	58.5	14364	14	US-10-067-443-20
16	725	58.5	14364	17	US-10-649-273-20
17	725	58.5	14364	17	US-10-651-722-20
18	468	37.7	371	18	US-10-430-201-3118
19	468	37.7	371	18	US-10-430-201-3119
20	337	27.2	1917	17	US-10-424-559-66417
21	270	21.8	1628	18	US-10-437-963-11249
22	257	20.7	1146	17	US-10-282-122A-11977
23	248.5	20.0	1000	18	US-10-343-561-50
24	248.5	20.0	1044	17	US-10-282-122A-26972
25	248.5	20.0	94750	17	US-10-672-787-38
26	247	19.9	936	17	US-10-282-122A-8315
27	241	19.4	1032	17	US-10-282-122A-11043
28	239	19.3	756	14	US-10-081-051-8
29	239	19.3	4360	14	US-10-081-051-2
30	231	18.6	1026	9	US-09-815-242-7701
31	231	18.6	1026	17	US-10-282-122A-30016
32	224	18.1	1029	9	US-09-815-242-6946
33	224	18.1	1029	17	US-10-282-122A-2020
34	224	18.1	1830121	17	US-10-329-670-1
35	224	18.1	1830121	18	US-10-158-865-1
36	213	17.2	1014	9	US-09-815-242-9682
37	213	17.2	1014	17	US-10-282-122A-39301
38	212	17.1	1020	17	US-10-282-122A-12254
39	210	16.9	1014	17	US-10-282-122A-11977
40	208	16.8	1007	17	US-10-282-122A-19220
41	208	16.8	1728	18	US-10-437-963-9137
42	207.5	16.7	1023	9	US-10-282-122A-11809
43	206	16.6	1014	17	US-09-815-242-6207
44	206	16.6	1014	17	US-10-282-122A-20511
45	206	16.6	1775	17	US-10-425-114-28614

ALIGNMENTS

RESULT 1
US-10-120-988-177
; Sequence 177, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyun
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802CON
; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_fl_genes Version 2.0

```
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
; US-10-988-177

Alignment Scores:
Pred. No.: 9,42e-152 Length: 1416
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-120-988-177 (1-1416)

QY 1 LeuEuaAlaEuaValGInGlyValSerAapPheLeuLeuGlyLeuSerLeuAaplle 20
DB CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 645
QY 21 AlaProGlyAapMetLeuAapLyValAlaArgArgLeuSerLeuIleYshIsProGlu 40
DB 646 GCACCAAGTGACATGCTTGACAGGTGGCAAGAACATTTCTTTATATAACATCCAGAG 705
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaYleGInGlyAsnArgPhe 60
DB 706 TGCTCCACCATGAGTGGTGGGAAACCATAGAACATTTGGCCAAACAGAAATAGATT 765
QY 61 HisPheAapIleYsProProLeuHisIshIsAlaYsAsnCysAapPheSerPheThrGly 80
DB 766 CATTTTGACATCAAACTCCCTTCATCATCATGCTAAATAATTTGATTTTCTTTTACGGA 825
QY 81 LeuGInHisValThrAapLySileIleMetLySlySGluYsGluGlyIleGluYs 100
DB 826 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGAGATTTGAGAG 885
QY 826 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGAGATTTGAGAG 885
QY 101 GlyCInIleLeuSerSerAlaAlaAapIleAlaIleValIleGInHisThrMetAlaCys 120
DB 886 GGGCAAAATCTGTCTTCAGACAGACATTCCTGCACAGTACACACATGGCATGT 945
QY 121 HisLeuVallysaArgThrHisArgAlaIleLeuPheCysYlySGInArgAapLeuPro 140
DB 946 CATTTTGAAAGAACACATCGGCTATCTGTTTGTAGACAGAGACTGTACTT 1005
QY 141 GInAsnAsnAlaValleuValAlaSerGlyValAlaSerAsnPheThrIleArgArg 160
DB 1006 CAAATAATATGCAATCTGTTGCAATCTGTCGTGTCGCAAGTAACTTTATATCCGCA 1065
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGInCysThrLeuLeuCysProProArgLeu 180
DB 1066 GCTCTGGAATTTTAAACAAACGACACACAGTGCATTTGTGTGTCTCTCCGACATA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1126 TGCACTGATTAATGCGATTATGATTCATGAGAAATGATTAAGAACTACGTGCTGCTTG 1185
QY 201 GlyIleLeuHisAapIleGluGlyIleArgGlyGluProLySProLeuGlyValAap 220
DB 1186 GGCATTTTACATGACATAGAAAGCATCCGCTATGAAACAAATATGCTCTTGGAATGAC 1245
QY 221 IleSerLySGluValGlyGluAlaSerIleYsValProGInLeuYsMetGluIle 239
DB 1246 ATATCAAAAGAGGTGGAGAGCTTCCATTAAGTAAACACATTAATAATGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TISSUE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-067-443-23

Alignment Scores:
Pred. No.: 1.05e-151 Length: 1526
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-23 (1-1526)

QY 1 LeuEuaAlaEuaValGInGlyValSerAapPheLeuLeuGlyLeuSerLeuAaplle 20
DB 85 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 144
QY 21 AlaProGlyAapMetLeuAapLyValAlaArgArgLeuSerLeuIleYshIsProGlu 40
DB 145 GCACCAAGTGACATGCTTGACAGGTGGCAAGAACATTTCTTTATATAACATCCAGAG 204
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaYleGInGlyAsnArgPhe 60
DB 205 TGCTCCACCATGAGTGGTGGGAAACCATAGAACATTTGGCCAAACAGAAATAGATT 264
QY 61 HisPheAapIleYsProProLeuHisIshIsAlaYsAsnCysAapPheSerPheThrGly 80
DB 265 CATTTTGACATCAAACTCCCTTCATCATCATGCTAAATAATTTGATTTTCTTTTACGGA 324
QY 81 LeuGInHisValThrAapLySileIleMetLySlySGluYsGluGlyIleGluYs 100
DB 325 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGAGATTTGAGAG 384
QY 101 GlyCInIleLeuSerSerAlaAlaAapIleAlaIleValIleGInHisThrMetAlaCys 120
DB 385 GGGCAAAATCTGTCTTCAGACAGACATTCCTGCACAGTACACACATGGCATGT 444
QY 121 HisLeuVallysaArgThrHisArgAlaIleLeuPheCysYlySGInArgAapLeuPro 140
DB 445 CATTTTGAAAGAACACATCGGCTATCTGTTTGTAGACAGAGACTGTACTT 504
QY 141 GInAsnAsnAlaValleuValAlaSerGlyValAlaSerAsnPheThrIleArgArg 160
DB 505 CAAATAATATGCAATCTGTTGCAATCTGTCGTGTCGCAAGTAACTTTATATCCGCA 564
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGInCysThrLeuLeuCysProProArgLeu 180
DB 565 GCTCTGGAATTTTAAACAAACGACACACAGTGCATTTGTGTGTCTCTCCGACATA 624
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 625 TGCACTGATTAATGCGATTATGATTCATGAGAAATGATTAAGAACTACGTGCTGCTTG 684
QY 201 GlyIleLeuHisAapIleGluGlyIleArgGlyGluProLySProLeuGlyValAap 220
DB 685 GGCATTTTACATGACATAGAAAGCATCCGCTATGAAACAAATATGCTCTTGGAATGAC 744
QY 221 IleSerLySGluValGlyGluAlaSerIleYsValProGInLeuYsMetGluIle 239
DB 744 ATATCAAAAGAGGTGGAGAGCTTCCATTAAGTAAACACATTAATAATGAGATA 1302
```

Db 745 ATATCAAAAGAGTTGAGAGCTTCATTAATAAGTACCAATTAATAATGGAGATA 801

RESULT 3

US-10-649-273-23

/ Sequence 23, Application US/10649273

/ Publication No. US20040043407A1

/ GENERAL INFORMATION:

/ APPLICANT: Bristol-Myers Squibb Company

/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

/ FILE REFERENCE: D0073 CNT

/ CURRENT APPLICATION NUMBER: US/10/649,273

/ CURRENT FILING DATE: 2003-08-27

/ PRIOR APPLICATION NUMBER: US 60/266,518

/ PRIOR FILING DATE: 2001-02-05

/ PRIOR APPLICATION NUMBER: US 10/067,443

/ PRIOR FILING DATE: 2002-02-05

/ PRIOR APPLICATION NUMBER: US 60/282,814

/ PRIOR FILING DATE: 2001-04-10

/ NUMBER OF SEQ ID NOS: 71

/ SOFTWARE: PatentIn version 3.2

/ SEQ ID NO 23

/ LENGTH: 1526

/ TYPE: DNA

/ ORGANISM: homo sapiens

US-10-649-273-23

Alignment Scores:

Pred. No.:	1.05e-151	Length:	1526
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-23 (1-1526)

QY 1 LeuLeuLaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db 85 CTGTGGCATTTAGTTCAAGAGATTTCAGATTTTCGCTTTGGAAAAGCTTTGGACATA 144

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

Db 145 GCACCAAGTGACATGCTTGAACCAAGTGCCAAAGAACTTTCTTTAATAAACATCCAGAG 204

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

Db 205 TGCCTCCCATGAGTGCGGAAAAGCCATAGAACATTTGGCCAAACAGAAATGATTT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80

Db 265 CATTTTGACATCAAACTCCCTTCATCATGCTAATAAATTTGATTTTCTTTTACTGGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGluGlyIleGluLys 100

Db 325 CTTCACACACCTTACTGATAATAATAAGAAAAAGAAAGAGGATTTGAGAG 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

Db 385 GGGCAAAATCCGTCTTCAAGCAGACATGCTGCAAGTACAGCACACAAATGGCATGT 444

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140

Db 445 CATCTTGTAAGAAAACACATCGGCTAATCTGTTTGTAAAGCAGAGAGACTTGTTACTT 504

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValValAlaSerAsnPheThrIleArgArg 160

Db 505 CAAATTAATAGCAGTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 564

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180

Db 565 GCTCTGGAATTTTAAACAAAGCAGACAGTCTTTGTTGTGCTCTCTCCAGACTA 624

QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200

Db 625 TGCACCTGATTAATGAGCATTAATGATTCATGAGAAATGGATGAAGACTACGCTGCTG 684

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgProLysCysAspProLeuGlyValAsp 220

Db 685 GGCATTTTATACATGACATGTAAGAGGCAATCCGCTATAGAACAAAATGCTCTTGAGTAGAC 744

QY 221 IleSerLysGluValGlyGluLysSerIleLysValProGlnLeuLysMetGluIle 239

Db 745 ATATCAAAAGAGTTGAGAGACTTCATTAATAAGTACCAATTAATAATGGAGATA 801

RESULT 4

US-10-651-722-23

/ Sequence 23, Application US/10651722

/ Publication No. US20040048302A1

/ GENERAL INFORMATION:

/ APPLICANT: Bristol-Myers Squibb Company

/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

/ FILE REFERENCE: D0073 DIV

/ CURRENT APPLICATION NUMBER: US/10/651,722

/ CURRENT FILING DATE: 2003-08-29

/ PRIOR APPLICATION NUMBER: US 60/266,518

/ PRIOR FILING DATE: 2001-02-05

/ PRIOR APPLICATION NUMBER: US 10/067,443

/ PRIOR FILING DATE: 2002-02-05

/ PRIOR APPLICATION NUMBER: US 60/282,814

/ NUMBER OF SEQ ID NOS: 71

/ SOFTWARE: PatentIn version 3.2

/ SEQ ID NO 23

/ LENGTH: 1526

/ TYPE: DNA

/ ORGANISM: homo sapiens

US-10-651-722-23

Alignment Scores:

Pred. No.:	1.05e-151	Length:	1526
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-23 (1-1526)

QY 1 LeuLeuLaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db 85 CTGTGGCATTTAGTTCAAGAGATTTCAGATTTTCGCTTTGGAAAAGCTTTGGACATA 144

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

Db 145 GCACCAAGTGACATGCTTGAACCAAGTGCCAAAGAACTTTCTTTAATAAACATCCAGAG 204

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

Db 205 TGCCTCCCATGAGTGCGGAAAAGCCATAGAACATTTGGCCAAACAGAAATGATTT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80

Db 265 CATTTTGACATCAAACTCCCTTCATCATGCTAATAAATTTGATTTTCTTTTACTGGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGluGlyIleGluLys 100

Db 325 CTTCACACACCTTACTGATAATAATAAGAAAAAGAAAGAGGATTTGAGAG 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

Db 385 GGGCAAAATCCGTCTTCAAGCAGACATGCTGCAAGTACAGCACACAAATGGCATGT 444

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140

Db 445 CATCTTGTAAGAAAACACATCGGCTAATCTGTTTGTAAAGCAGAGAGACTTTTACTT 504

[illegible]

```

RESULT 5
US-10-067-443-1
; Sequence 1, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEINASE HIGHLY EXPRESSED
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-067-443-1

```

Alignment Scores:	
Pred. No.:	1,79e-151
Score:	1240.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
BB:	14
Length:	2197
Matches:	239
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) X US-10-067-443-1 (1-2197)

QY	1	LeuLeuAlaLeuValGIGlYValSerAspPheLeuLeuLeuGlyysSerLeuAspIle	20
Db	756	CTGTGGCATTAAGTTCACAGAGCTTTCAGATTTTCTGCTCTTGAAAGTCTTTGGACATA	815
QY	21	AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysIAspProGlu	40
Db	816	GCACACAGGTGACATCTTGACCAAGGCGCAAGAGACTTCTTTAATTAACATCCAGAG	875
QY	41	CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAspAsnArgPhe	60
Db	876	TGCTTCACCATGAGTGCGTGGGAAGCCATAGAAcATTGGCCCAACACAGGAATTAAGATTT	935
QY	61	HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	80
Db	936	CATTTCACATCAACACCTCCCTTGCAATCACTCAAAAATTTTGATATTTTCTTTTACACGA	995
QY	81	LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGluGlyIleGluLys	100

Dd	996	CTTCAACGCTTACTGATTAATAATATGAAAAAGAAAGAAAGATTTGAGAG	1055
Qy	101	GLYGINILEUUSERSERIALAASPILEALATHRYALGINHSTRIMETALCY	120
Dd	1056	GAGCAAAATCCTGTCTTCGACGACAGACATTGCTCCACAGTACAGACACAATGGCAGT	1115
Qy	121	HISLEUVALYBAGTHHSATGALALILEUPHCEYBLYSGINADGASPILEUENPRO	140
Dd	1116	CATCTTGTGAAAAGAACACATCGGGCAATTCGTGTTGTATGACGAGAGCATTTGTACT	1175
Qy	141	GLINASPMANALAVALLEUVALIASERGLYGLYVALIASERASPMETRYLLEARGARG	160
Dd	1176	CAAAATATATGCAAGTACGTGTGCAATCGTGCTGCCAATTAATCTTATATCCGACGA	1235
Qy	161	ALALEUGNILEUSTRANASALATHRGINCYSTRILEUENCYSPROPROARGLEU	180
Dd	1236	GCTCTGGAAATTTTAAACAAAGCAACACAGGCACTTTGTTGTCTCTCCACAGACTA	1295
Qy	181	CYSTRHAPASANGIYILEMERILEALATRYANGLYILEGUARGLEUARGALIGLYLEU	200
Dd	1296	TGCACGTATATATGCAATTATGATTCGATGGAATGGTATTGAAAAGACTACGGCTGCCTTG	1355
Qy	201	GLYILEUENHISASPILEGLUGLYLLEARGTRYGUNPOLYCYSPROLEUGLYVALASP	220
Dd	1356	GGCAATTTTACATGACATAGAGGAGCATCCGCTATATACCAAAATGTCCTCTTGGAGTAGAC	1415
Qy	221	ILESERLYSGIVAGLYGVALASERILEYVALPROGINLEUYSMETGINILE	239
Dd	1416	ATATCAAAAAGAGTTGAGAAAGCTTCCATTAATAAGTACCAATTTAAAATGGAAGATA	1472

RESULT 6
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1

```

? APPLICANT Bristol-Myers Squibb Company
? TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL MENTHOLPROTEASE, MP-1
? FILE REFERENCE: D0073 CNT
? CURRENT APPLICATION NUMBER: US/10/649,273
? CURRENT FILING DATE: 2003-08-27
? PRIOR APPLICATION NUMBER: US 60/266,518
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 10/667,443
? PRIOR FILING DATE: 2002-02-05
? PRIOR APPLICATION NUMBER: US 60/282,814
? PRIOR FILING DATE: 2001-04-10
? NUMBER OF SEQ ID NOS: 71
? SOFTWARE: PatentIn version 3.2

```

LOCATION: (231) .. (1472)

Alignment Scores:

Pred. No.:	1,79e-151	Length:	2197
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-1 (1-2197)

Qy 1 LeuLeuAlaIeuValGlnGlyValSerAspPheLeuLeuGlyValysSerLeuAspIle 20
Db 756 CTGTGGCATTAGTTCAGAGACTTTTCTGCTCTTGGAAGCTTTGGACATA 81
Qy 21 AlaProGlyAspMetLeuAspIysValAlaIaArgLeuSerLeuIleLysHisProGlu 40

Db	Seq ID	Accession	Length	Score	Matches
Db	816	GCACAGGTGACATCTTGACAGAGGTGGCAAGAAACATTTCTTTATATAAACATCCAG	875	1.79e-151	2197
Qy	41	CysSerThrMetSerGlyGlyValaIleGluHisLeuAlaValGlnGlyAsnArgPhe	60	1.79e-151	2197
Db	876	TGCTCCACCAATGAGTGTGGGAAACCATAGAACATTTGGCCAAACAGAAATGATTT	935	1.79e-151	2197
Qy	61	HisPheAspIleValProProLeuHisHisAlaValAsnGlyAsnPheSerPheThrCly	80	1.79e-151	2197
Db	936	CATTTTGACATCAAACTCTCCCTGATCAATGCTAAATAATGTGAATTTTCTTTTACGGA	995	1.79e-151	2197
Qy	81	LeuGlnHisValaThrAspLysIleIleMetLysGlyGluValGluGluGlyIleGluLys	100	1.79e-151	2197
Db	996	CTTCACACACGCTTACGATTAATTAATTAATGAAGAAAGGAAAGGAGATTTGAGAG	1055	1.79e-151	2197
Qy	101	GlyGlnIleLeuSerSerAlaIleAspIleAlaIleValaIleValaIleGlnHisThrMetAlaCys	120	1.79e-151	2197
Db	1056	GGGCAAAATCCCTGCTTCCAGACAGACATTCCTGCCACAGTACAGACACAAATGGCATGT	1115	1.79e-151	2197
Qy	121	HisLeuValaLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro	140	1.79e-151	2197
Db	1116	CATCTTGAAAGAAACACATCGGGCTATTCGTTTGTAAAGCAGAGACTGTGTACCT	1175	1.79e-151	2197
Qy	141	GlnAsnAsnAlaValaIleValaIleAsnSerGlyValaIleSerAsnPheTyrIleArgArg	160	1.79e-151	2197
Db	1176	CAAAATATATGACATGATCTGTTGATCTGCTGTGTCGAGTAACTTTATATTCGACGA	1235	1.79e-151	2197
Qy	161	AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	180	1.79e-151	2197
Db	1236	GCTCGAAATTTTAAACAAACGACACACAGTGCATCTTTGTTGTCCTCCACGACTA	1295	1.79e-151	2197
Qy	181	CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluValGluLeuAlaGlyLeu	200	1.79e-151	2197
Db	1296	TGCATGTATTAATGGCATTAATATGATGATGATGAAATGGTATTAACGCTGCTGGCTTG	1355	1.79e-151	2197
Qy	201	GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValaAsp	220	1.79e-151	2197
Db	1356	GGCATTTTACATGACATAGAAAGGATCCGCTATGAAACAAATATGCTCTTGAGATGAC	1415	1.79e-151	2197
Qy	221	IleSerLysGluValaGlyValaAsnIleLysValaProGlnLeuLysMetGluIle	239	1.79e-151	2197
Db	1416	ATATCAAAAGAAAGTGGAGAAAGCTTCATTAATAAGTACCAATTAATAATGAGATA	1472	1.79e-151	2197

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Missmatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-1 (1-2197)

QY 1 LeuLauAlaLeuValGInGlyValSerAspPheLeuLeuGlyIlySerLeuAspIle 20
DB 756 CTGTTGGCATTGATGTCACAGAGATTTCAGATTTCTGCTTCCTTGAAAGCTCTTGGAACATA 815
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHieProGlu 40
DB 816 GCAACAGGAGCAATGCTTGAACAGGTGGCAAGAAAGACTTCTTTATATAAACATCCAGAG 875
QY 41 CysSerThrMetSerGlyIlyLysAlaIleGluHieLeuAlaLysGlnGlyAsnArgPhe 60
DB 876 TGCTCCACCATGATGATGTTGGAAAGCCATAGAACATTGGCCAAACAGAAATATAGATT 935
QY 61 HisPheAspIleLysProProLeuHieHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 936 CATTGTGACATGAACCTCTCCCTTCATCATGCTAAATAATGTGATTTTCTTTTACATGGA 995
QY 81 LeuGlnHisValThrAspLysIleIleMetLysIleGluLysGlnGluGlyIleGluLys 100
DB 996 CTTCAACACGTTACTGATTAATATATATGAAAGAAAGAAAGAAAGATATTGAGAG 1055
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1056 GGGCAAACTCTGCTCTTCAACAGACAGCATTTGCTGCCACATGACACACAAATGCAATGT 1115
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1116 CATCTTGGAAGAAAGAACACATCGGGCTATTCTGTTTGTGAAGCAGAGACTTGTTACT 1175
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyIlyValAlaSerAsnPheTyrlleArgArg 160
DB 1176 CAATATATATGACGATCTGATGCTGATGCTGGTGGTGGCAAGTATCTTATATCCGACA 1235
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1236 GCTCGGAAATTTTAAACAAGCAACACAGTGCACTTGTGTGCTCTCCACAGACTA 1295
QY 181 CysThrAspAsnGlyIlyIleMetIleAlaTrpAsnGlyIlyLeuArgLeuAlaGlyLeu 200
DB 1296 TGCATCTGATTAAGGCAATTATGATTCATGCAATGATGATTAAGAAACATACGTGCGCTTG 1355
QY 201 GlyIleLeuHisAspIleGluGlyIlyleArgTyrlleGluProLysCysProLeuGlyValAsp 220
DB 1356 GGCAATTTTAAACAGACATGAAGGCAATCCGCTATGAACCAAAAGTCTCTTGAGTGAAC 1415
QY 221 IleSerLysGlyValGlyGluLysSerIleLysValProGlnLeuLysMetGluIle 239
DB 1416 ATATCAAAAGAAAGTGGAGAACCTTCCATTAAGTACCAATTAATAATGAGATA 1472

RESULT 8
US-10-067-443-21
: Sequence 21, Application US/10067443
: Publication No. US20030082782A1
: GENERAL INFORMATION:
: APPLICANT: Bristol-Myers Squibb Company
: TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
: FILE REFERENCE: D0073 NP
: CURRENT APPLICATION NUMBER: US/10/067,443
: CURRENT FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US 60/266,518
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 60/292,914
: PRIOR FILING DATE: 2001-04-10
: NUMBER OF SEQ ID NOS: 71
: SOFTWARE: Patent version 3.0
: SEQ ID NO 21

```


; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:

Pred. No.:	2,72e-148	Length:	1387
Score:	1213.50	Matches:	238
Percent Similarity:	90.15%	Conservative:	0
Best Local Similarity:	90.15%	Mismatches:	1
Query Match:	97.86%	Indels:	25
DB:	14	Gaps:	1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-21 (1-1387)

```
QY      1 LeuLeuAlaLeuValGlnGlyValSerApphLeuLeuGlyLysSerLeuAspLe 20
DB      549 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGGAAGCTTTGACATA 608
QY      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB      609 GCACCAAGTGACATGCTTGACAGAGTGCGCAAGACATTTCTTTAATTAACATCCAGAG 668
QY      41 CysSerThMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB      669 TGCTCCACCATGAGTGGTGGGAAAGCCATAGAGCATTTGGCCAAACAAGAAATGATTT 728
QY      61 HisPheAspLysLeuProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB      729 CATTGTGACATCAAACTCCCTTGATCATGTCTAAATAATGATTTTCTTTTACTGGA 788
QY      81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGlyIle----- 98
DB      789 CTTCAACACGTTACTGATTAATAATTAATGAAGAAAGAAAGAGAGATATATTCTTA 848
QY      98 ----- 98
DB      849 ATTAGTAAGTTGAACAGATAAATATTCCTGATTTGCTCTAAATAAGCTGCTATTTC 908
QY      99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
DB      909 TGCAGGTATGAGAAAGGGGCAAAATCCTGCTTTCAGCAGCAGACATTTGCGCACAGTACAG 968
QY      116 HisThrMetAlaCyHisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB      969 CACCAATGGCATGTCTCTTGAAAGAAACAACATCGGGCTAATCTGTTTGTAGACAG 1028
QY      136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaIleSerAsn 155
DB      1029 AGAGACTTGTACTCTCAAAATTAATGCAATCTGCTTGCATCTGCTGCTGCTGCAAGTAAAC 1088
QY      156 PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175
DB      1089 TTCTATATCCGCAAGAGCTCTGGAATTTTAAACAACGCAACAGTGCACTTTGTTGTGT 1148
QY      176 ProProProAlaGluLeuCysThrAspAsnGlyIleMetIleAlaIleTrpAsnGlyIleGluArg 195
DB      1149 CCTCTCCCAACATGACATGATATGCAATTAATGATTAATGATTAATGATTAATGATTA 1208
QY      196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCys 215
DB      1209 CTAGCTGTGCTGCTGGGCAATTTTACATACATAGAAAGGATCGCTAGAACCAAAATGT 1268
QY      216 ProLeuGlyValAspLysSerLysGlyValGlyGlnAlaSerIleLysValProGlnLeu 235
DB      1269 CCTCTTGGAGTAGACATATCAAAAGAGAGTGGAGAGAGCTTCATAAAGTATCACACATTA 1328
QY      236 LysMetGluIle 239
DB      1329 AAAATGAGATA 1340
```

RESULT 9
US-10-649-273-21

```
; Sequence 21, Application US/10649273  
; Publication No. US2004003407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-649-273-21
```

Alignment Scores:

Pred. No.:	2,72e-148	Length:	1387
Score:	1213.50	Matches:	238
Percent Similarity:	90.15%	Conservative:	0
Best Local Similarity:	90.15%	Mismatches:	1
Query Match:	97.86%	Indels:	25
DB:	17	Gaps:	1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-21 (1-1387)

```
QY      1 LeuLeuAlaLeuValGlnGlyValSerApphLeuLeuGlyLysSerLeuAspLe 20
DB      549 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGGAAGCTTTGACATA 608
QY      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB      609 GCACCAAGTGACATGCTTGACAGAGTGCGCAAGACATTTCTTTAATTAACATCCAGAG 668
QY      41 CysSerThMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB      669 TGCTCCACCATGAGTGGTGGGAAAGCCATAGAGCATTTGGCCAAACAAGAAATGATTT 728
QY      61 HisPheAspLysLeuProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB      729 CATTGTGACATCAAACTCCCTTGATCATGTCTAAATAATGATTTTCTTTTACTGGA 788
QY      81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGlyIle----- 98
DB      789 CTTCAACACGTTACTGATTAATAATTAATGAAGAAAGAAAGAGATATATTCTTA 848
QY      98 ----- 98
DB      849 ATTAGTAAGTTGAACAGATAAATATTCCTGATTTGCTCTAAATAAGCTGCTATTTC 908
QY      99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
DB      909 TGCAGGTATGAGAAAGGGGCAAAATCCTGCTTTCAGCAGCAGACATTTGCGCACAGTACAG 968
QY      116 HisThrMetAlaCyHisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB      969 CACCAATGGCATGTCTCTTGAAAGAAACAACATCGGGCTAATCTGTTTGTAGACAG 1028
QY      136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaIleSerAsn 155
DB      1029 AGAGACTTGTACTCTCAAAATTAATGCAATCTGCTTGCATCTGCTGCTGCTGCAAGTAAAC 1088
QY      156 PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175
DB      1089 TTCTATATCCGCAAGAGCTCTGGAATTTTAAACAACGCAACAGTGCACTTTGTTGTGT 1148
QY      176 ProProProAlaGluLeuCysThrAspAsnGlyIleMetIleAlaIleTrpAsnGlyIleGluArg 195
```

Accession	Protein	Length (aa)
D8	1149 CCTCTCCCAACATATGCACTGATPATGGCATTTGATTCGATGGAAATGGATTTGAAGA	1208
QY	196 LeuArlgaLaGIleuGIYIleuHIsApIleGIuGIYleArGYrGIuProIeCyS	215
D8	1209 CTAGTGGTGGCTTGGGCACTTTTACATACATAGAGGCACTCCGTATGAACCAAAAGT	1268
QY	216 ProIeuGIyLaApIleSerIeSGIuValGIyGuaIaSerIleYsValProGIuDeu	235
D8	1269 CCTCTGGAGTACATATCAAAAGAACTTGAGAGACTTCATATAAGTACCAATTA	1328
QY	236 LysMetGIuIle 239	
D8	1329 AAAATGGAGATA 1340	

RESULT 10

```

? Sequence 21, Application US/10651722
? Publication No. US2004004830A1
? GENERAL INFORMATION:
? APPLICANT: Bristol-Myers Squibb Company
? TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
? FILE REFERENCE: D0073 DIV
? CURRENT APPLICATION NUMBER: US/10/651,722
? CURRENT FILING DATE: 2003-08-29
? PRIOR APPLICATION NUMBER: US 60/566,518
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 10/067,443
? PRIOR FILING DATE: 2002-02-05
? PRIOR APPLICATION NUMBER: US 60/282,814
? PRIOR FILING DATE: 2001-04-10
? NUMBER OF SEQ ID NOS: 71
? SOFTWARE: Patentin version 3.2
? SEQ ID NO 21
? LENGTH: 1387
? TYPE: DNA
? ORGANISM: homo sapiens
? US-10-651-722-21

```

US-10-651-722-21

Alignment Scores:	2.72e-148	Length:	1387
Pred. No.:		Matches:	238
Score:	1213.50	Conservative:	0
Percent Similarity:	90.154	Mismatches:	1
Best Local Similarity:	90.154	Indels:	25
Query Match:	97.864	Gaps:	1
DB:	17		

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-21 (1-1387)

Qy	1	LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIle	20
Db	549	CTGTGGCATTAAGTTCAAGAGATTTCAGATTTTCTCTCTCTTGGAAAGCTTTTGGACATA	600
Qy	21	AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu	40
Db	609	GCACACAGGTGACATCTCTTGACAAAGGGCGCAAGAGCTTTCCTTAATTAACATCCACAG	668
Qy	41	CysSerTrpMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe	60
Db	669	TGCTCCACATGAGTGTGGGAAAGCCATAGAGCATTTGGCCCAACAGAAATAGATTTT	728
Qy	61	HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	80
Db	729	CATTGTGACATCAACCTCTCCCTGTCATCTAGCTAAATAATGTGATATTTTCTTTTACCTGA	788
Qy	81	LeuGlnHisValAlaThrAspLysIleIleMetLysGlnLysGlnGlyIle-----	98
Db	789	CTTCACACGTTACTGATTAATTAATATGAAAGAAAGAGAGAGGTATATTTCTTA	848
Qy	98	-----	98

Oy	99	-----GlytSGiYgInIleLeuSerSerAlaIAspIleAlaIaThValGIn	115
Db	909	TGCGAGTATGAAAGGGGCMAATCTCTTCAAGACAGACATTGCTCCACAGTAC	968
Oy	116	HisThMeAlaCyShiSLeuValIysArgThrHisArgAlaIleLeuPheCyValSGIn	135
Db	969	CACACAAATGGCATCTCATCTTGTAAGAAACACATCGGGGTATTCTGTTTGTAAACAG	1028
Oy	136	ArgAspLeuPProGInaAspAsnAlaValLeuValIaSerGlyValAlaSerAsn	155
Db	1029	AGAACTTGTAACCTCAAAATAAATGACAGTACTGTGTCATCTGGTGGTGGCAGATAC	1088
Oy	156	PheTrIleArgArgAlaLeuGulIleLeuThrAsnAlaThrGInCySthrLeuLeuCyS	175
Db	1089	TTCTATATCCGCAAGAGCTGTGAAATTTTAAACAACCAACACAGTCACATTGTTGTGT	1148
Oy	176	ProProProAlaGLeuCySThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArg	195
Db	1149	CCCTCTCCCAAGACTATACATCGAATAAATGGCATTTATGATTCGATGGAAATGGTATTTGAAGA	1208
Oy	196	LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTrpGluProLeuCyS	215
Db	1209	CTACGTGTGGCTTGGGGCATTTTACATGACACTAGAAAGCATCCGCTATGAACCAAAATGT	1268
Oy	216	ProLeuGlyValAspIleSerIysGluValGlyGluIaSerIleIysValProGInLeu	235
Db	1269	CCCTCTGGAGTAGACATATCAAAAGAAGTTGGAGAACCTTCATPAAAAAGTACCAATTA	1328
Oy	236	LysMetGluIle	239
Db	1329	AAATGGAGATA	1340

RESULT 11

; Sequence 6, Application US/10012140
; Publication No. US20030009017A1

GENERAL INFORMATION:

```

1  APPLICANT: Leidy, Kevin R.
2  APPLICANT: Kapeller-Libermann, Rosana
3  APPLICANT: Glucksmann, Maria A.
4  TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
5  TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
6  TITLE OF INVENTION: THERBOF
7  FILE REFERENCE: 381552004900
8  CURRENT APPLICATION NUMBER: US/10/012,140
9  CURRENT FILING DATE: 2001-11-08
10 PRIOR APPLICATION NUMBER: 60/246,768
11 PRIOR FILING DATE: 2000-11-08
12 PRIOR APPLICATION NUMBER: 60/246,772
13 PRIOR FILING DATE: 2000-11-08
14 PRIOR APPLICATION NUMBER: 60/249,185
15 PRIOR FILING DATE: 2000-11-15
16 NUMBER OF SEQ ID NOS: 49
17 SOFTWARE: FastSeq for Windows Version 4.0
18 SEQ ID NO 6
19 LENGTH: 1245
20 TYPE: DNA
21 ORGANISM: Homo sapiens
22 US-10-012-140-6

```

Alignment Scores:

Pred. No.:	5.52e-147	Length:	1245
Score:	1203.00	Matches:	232
Percent Similarity:	98.33%	Conservative:	3
Best Local Similarity:	97.07%	Mismatches:	4
Query Match:	97.02%	Indels:	0
DB:	14	Gaps:	0

US-10-649-273-2 COPY 176 414 (1-229) X US-10-012-140-6 (1-1245)

Oy 1 LeuleuaIaleuValGInglYvalSerAspPheleuleuLeuCllyVserleuAspile 20
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 526 CTGTGGCAATTAGTTCAGAAGTTCAAGATTTCTGCTTCTTGAAACTCTTGACATA 585

QY	21	AlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleYrsHisAspGlu	40
Db	586	GCACCAAGGTGACATCTCTTCAACAGGTGGCAAGAGCTTTCTTTAATAAATCCAGAG	645
QY	41	CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnIleAsnArgPhe	60
Db	646	TGCTCCACCATATGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT	705
QY	61	HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	80
Db	706	CATTGTGACATCAAACTCTCCCTTGATCATAGCTCAAAAAATTGGATTCTTTTACTGGA	765
QY	81	LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys	100
Db	766	CTTCACACGCTTACTGATTAATAATAATAGAAACAGAAACAGAAACAGATATTGGAAG	825
QY	101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys	120
Db	826	GGGCAAAATCTGTCTTCCAGCAGACAGATTCGTCGACAGTACAGACACAATGCGATGT	885
QY	121	HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro	140
Db	886	CATCTTGGAAAAAGAACACATCGGGCTATCTGTTTGTAGACAGAGAACCTTGTAACCT	945
QY	141	GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg	160
Db	946	CAAAATAATGACGATCTGGTTGCATCTGGTGGGTGCGCAGATGACTTCTATATCCGACA	1005
QY	161	AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	180
Db	1006	GCTCGGAAATTTTAAACAAACGCAACACAGTGCATCTTGTGTGCTCTCCCAACTA	1065
QY	181	CysThrAspAsnGlyIleMetIleAlaThrPasnGlyIleGluArgLeuArgAlaGlyLeu	200
Db	1066	TGCATCTGATTAAGGCACTTATATGATGATGATAATGATTAAGAAAGACTACGTGCTGCTTG	1125
QY	201	GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp	220
Db	1126	GGCATTTTACATGACATAGTAAGAGGATCCGCTATGAAACCAAAATGCTCTTGGAGATGAGC	1185
QY	221	IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle	239
Db	1186	ATATCAAAAGAGTGGAGAGACCTTCATATAAAGTACCAATTAATAAATGAGAGATA	1242

RESULT 12

US-10-012-140-4

Sequence 4, Application US/10012140

Publication No. US20030009017A1

GENERAL INFORMATION:

APPLICANT: Leiby, Kevin R.

APPLICANT: Kapeller-Libermann, Rosana

APPLICANT: Glucksmann, Maria A.

TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES

TITLE OF INVENTION: THEREOF

FILE REFERENCE: 381552004900

CURRENT FILING DATE: US/10/012,140

PRIOR FILING DATE: 2001-11-08

PRIOR APPLICATION NUMBER: 60/246,768

PRIOR FILING DATE: 2000-11-08

PRIOR APPLICATION NUMBER: 60/246,772

PRIOR FILING DATE: 2000-11-08

PRIOR APPLICATION NUMBER: 60/249,185

NUMBER OF SEQ ID NOS: 49

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 1820

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

[illegible]

```

; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKI, ICHIRO
; APPLICANT: SEKI, NAOTIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-400

```

```

Alignment Scores:
Pred. No.: 9,58e-128 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 17 Gaps: 1

```

US-10-649-273-2_COPY_176_414 (1-239) x US-10-094-749-400 (1-2208)

```

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
DB 869 CTGTTGGCATTAAGTTCAAGAGATTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 928
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 929 GCACCAAGTGACATCTTGACAAAGTGCGCAAGAGCTTCTTAATAAACAATCCAGAG 988
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 989 TGCTTCACCATGACGTGTGGGAAAGCCATAGAACTTTGGCCAAACAGAAATATGATTT 1048
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 1049 CATTTTGACATCAACCTCCCTTGATCATAGCTAAATAATGATTTCTTTTACTGGA 1108
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLys 100
DB 1109 CTTCAACACGTTACGATTAATAATTAATGAAGAAAGAAAGAGAGATTAAGAGAG 1168
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleArgThrValGlnHisThrMetAlaCys 120
DB 1169 GGGCAAAATCTGTCTTCAAGCAGACACATTCCTGCACAGTACAGCACAAATGCGATGT 1228
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 1229 CATCTTGAAAGAAACACATCGGCTATCTGTTTGAACAGAGACATTTGTAACCT 1288
QY 141 GlnSerAsnAlaValIleuValAlaSerGlyValAlaLysAsnPheTyrIleArgArg 160
DB 1289 CAAATATATGCACTGTGCTTGCATCTGTGTGCTGCGAGTACTTCTGTATCCGACA 1348

```

```

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1349 GCTCTGGAATTTTAACAACGACACAGTGACCTTTGTGTCTCTCCCAAGACTA 1408
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
DB 1409 TGCACGTATATGCGCATTAATGATTTGCA----- 1435
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProLysCysProLeuGlnValAsp 220
DB 1436 -----TGATGTCTCTTGGAGTAGAC 1456
QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239
DB 1457 ATATCAAAAGAGCTTGAGAGAGCTTCCATTAAGTACCAATTAATAAAGAGATTA 1513

```

```

RESULT 14
US-10-723-860-7447
; Sequence 7447, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nacasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7447
; LENGTH: 2890
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (646)..(657)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-723-860-7447

```

```

Alignment Scores:
Pred. No.: 1.42e-127 Length: 2890
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 18 Gaps: 1

```

US-10-649-273-2_COPY_176_414 (1-239) x US-10-723-860-7447 (1-2890)

```

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
DB 1526 CTGTTGGCATTAAGTTCAAGAGATTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 1585
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 1586 GCACCAAGTGACATCTTGACAAAGTGCGCAAGAGCTTCTTAATAAACAATCCAGAG 1645
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 1646 TGCTTCACCATGACGTGTGGGAAAGCCATAGAACTTTGGCCAAACAGAAATATGATTT 1705
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 1706 CATTTTGACATCAACCTCCCTTGATCATAGCTAAATAATGATTTCTTTTACTGGA 1765
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLys 100
DB 1766 CTTCAACACGTTACTAATAATTAATGAAGAAAGAAAGAGATTAAGAGAG 1825

```

QY 101 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1826 GGGCAATCTCTCTTCAGCAGCAGACATCTGTCACAGTACAGCAGCAATGGCATGT 1885
QY 121 HisLeuValIleAspThrHisArgAlaIleLeuPheCysIleGlnArgAspLeuPro 140
DB 1886 CATCTTGGAAGAAACACATCGGCTATTCTGTTTGTAGCAGAGACATTTGTAACCT 1945
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArgArg 160
DB 1946 CAAATTAATGACAGTCTGTCATCTGTCAGTGTGCGAAGTAACCTTGTAATCCGCA 2005
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 2006 GCTCGAATATTTTAACAAACGCAACAGTGCATTTGTTGTCTCTCCAGACTA 2065
QY 181 CysThrAspAspGlyIleMetIleAlaTrpAspGlyIleGlnArgLeuArgAlaGlyLeu 200
DB 2066 TGCACTGATTAATGCGATTATGATGCA----- 2092
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 2093 -----TGATGTCCTCTTGAGATGAC 2113
QY 221 IleSerIleGluValGlyIleValAspIleLeuValProGlnLeuIleMetGluIle 239
DB 2114 ATATCAAAAGAAAGTTGGAGAGCTTCCATTAAGTACCAATTAAATGAGATTA 2170

RESULT 15

US-10-067-443-20
: Sequence 20, Application US/10067443
: Publication No. US20030082782A1
GENERAL INFORMATION:
: APPLICANT: Bristol-Myers Squibb Company
: TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEINASE HIGHLY EXPRESSED IN
: FILE REFERENCE: D0073 NP
: CURRENT APPLICATION NUMBER: US/10/067,443
: PRIOR FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US 60/266,518
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 60/282,814
: NUMBER OF SEQ ID NOS: 71
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 20
: LENGTH: 14364
: TYPE: DNA
: ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:

Pred. No.: 8,936-83 Length: 14364
Score: 725.00 Matches: 186
Percent Similarity: 32.86% Conservatve: 0
Best Local Similarity: 32.86% Mismatches: 0
Query Match: 58.47% Indels: 380
DB: 14 Gaps: 3

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-20 (1-14364)

QY 29 ValAlaAlaGluGluSerLeuIleLeuHisProGluCysSerThrMetSerGlyGlyIleVal 48
DB 11840 GTGGCAAAAGACTTTCTTTTATTAACATCCAGAGTGTCCACCAATGAGTGTGGGAAA 11899
QY 49 AlaIleGlnHisLeuAlaValGlnGlyAsnArgPheHisAspAspIleLysProProLeu 68
DB 11900 GCCATGAAACATTGGCCAAAGAAAGAAATGATTTGATTTGACATCAACCTCCCTTG 11959
QY 69 HisHisAlaIleValAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIle 88
DB 11960 CATCATGCTAAATAATGTGATTTTCTTTTACTGACATTCACAGCTTACTGATAAATA 12019

QY 89 IleMetIleValGlnIleValGlnGlu----- 96
DB 12020 ATTAATGAAAAAGAAAAAGAGAGATATTTCTTAATTAGTAAGTGAACAGATAAAT 12079
QY 97 -----GlyIleGlnIleValGlnIle 103
DB 12080 ATTCCTGATTTGCTTAATAAATAGCTGCTATTCTTCAGAGGATTAAGAAAGGGCAAAAT 12139
QY 103 IleuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCysHisIleVal 123
DB 12140 CCGTCTTCAGCAGCAGCAATTTGCTGCCACAGTACAGCACAATGCAATGCTATCTTG 12199
QY 123 IlyAspArgThrHisArgAlaIleLeuPheCysIleGlnArgAspLeuProGlnAsnAs 143
DB 12200 GAAAAAGAACATCGGGCTATTCTGTTTGTAAAGCAGAGAGACTTGTACTCAAAATAA 12259
QY 143 nAlaValLeu----- 146
DB 12260 TGCACTGATGTAAGTTTATCTATTATAGTAATAGTTACACTTGCAATATGTAC 12319
QY 146 ----- 146
DB 12320 TTTTTCACAGACCTTGAACCTTGCTGTAGATGAGTAAGACATCTTATGCTTATGCTAG 12379
QY 146 ----- 146
DB 12380 CCTGACAGTATGAATTTATGACAGATAGAGAAAGACTAAACGCCATTTCTGTACTAGTT 12439
QY 146 ----- 146
DB 12440 TGGTAGCTTATAGGACAGCTGATATAGCTTATATGACACATAAGTCTAAATTTGACATTT 12499
QY 146 ----- 146
DB 12500 CTGTGTGATTTAAAGAGGGCTTACAAATTAAGAAAGTAATGACAGTAACGTATACACT 12559
QY 146 ----- 146
DB 12560 AATTTTAAAGAAATAGTGAGATTTCTTCACTCTTGATGAATCCCTTTGTTGTTGT 12619
QY 146 ----- 146
DB 12620 TTTTAAATAGCCAGTCAATTTTACAGATGGAGGTGTATTTCAACTTGTGTACACT 12679
QY 146 ----- 146
DB 12680 AATGTGATTAAGTCTGTATTAATCCATTAATTTGACAGCCAAATCCCTTAATTTGTG 12739
QY 146 ----- 146
DB 12740 CTTAAAGCCTTGACAAACATCCTGTTTAATCTGTATCTTAACCTTAATTAATAAAT 12799
QY 146 ----- 146
DB 12800 TATTAACATAAGTGGAAATGTTTAAATGTAGTAATTCATAGATGAATTTTACATGG 12859
QY 146 ----- 146
DB 12860 ATATCAAGAAATATTTTTCAGAGTATGTAGTAATAAGCAAAATATATAAATTTTC 12919
QY 146 ----- 146
DB 12920 AGGCTTAATAATAGTGTACTATGATTAATAATTAATAATTAATTAATAGATAAGGT 12979
QY 146 ----- 146
DB 12980 TGGAGAAATAATATCAAAATGCTAGTAATGTTGTATGCTATTAAGATTTATAGTAAT 13039
QY 146 ----- 146
DB 13040 TTTTCTTCAAAATTTTATATACATAGATATGTATGCTGCCATTAACCATCTCAAAAT 13099
QY 146 ----- 146

```
Db 13100 GGGATAGTTATTATTGTTTAATGCTGATATTTTCTCAGGTTTAAATAGCAGCTTGGT 13159
QY 147 -----ValAlaSerGlyG1 151
Db 13160 TCATATCCATATATGATATGATATTTTGGTTTCTCATTCCTTCAGGTTCATCTGGTGG 13219
QY 151 ValAlaSerAsnPhenylleArgArgAlaLeuGluileuThrAsnAlaThrgInCy 171
Db 13220 TGTGGCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTAACAAACGCAACACAGTG 13279
QY 171 sThrLeuLeuCyProProArgLeuCyThrAspAsnGlyilemeCileAlaTrp-- 190
Db 13280 CACTTGTGTGTCTCTCCAGACTATGCACTGATATGCAATATGATGATGCAATGTA 13339
QY 190 ----- 190
Db 13340 AGCCACAGATATATAGTCTTCACTCATTAATGTAATTAATGCAATTTATTCAT 13399
QY 190 ----- 190
Db 13400 ACTAAGCTTCTCTCTCAGATCTTGAGCTATTTGATTTTAATTTTAAGCTTCTTATTTA 13459
QY 191 --AsnGlyileGluArgLeuArgAlaGlyleuGlyileuHisAspIleGluGlyileA 210
Db 13460 GGAATGGATTTGAAGAAGACTAGGTGCTGGCTTGGCATTTTACATGACATAGAAAGCATCC 13519
QY 210 rglYrGluProlys 214
Db 13520 GCTATGAACCAAG 13533
```

Search completed: February 17, 2005, 01:27:31
Job time : 500 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:38:06 ; Search time 10.9109 Seconds
(without alignments)
2107.605 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVQGVDFLLGKSLDI.....DISKVEGASIKVPLQKMEI 239

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	326	26.3	463	2	B84888
2	244	19.7	365	2	A82902
3	244	19.7	366	2	D97677
4	241.5	19.5	359	2	AB3274
5	238	19.2	387	2	E71711
6	237.5	19.2	367	2	F87257
7	231	18.6	341	2	H83572
8	230.5	18.6	344	2	E97707
9	224	18.1	342	2	H64074
10	221.5	17.9	323	2	G69388
11	213	17.2	337	2	AG0892
12	210.5	17.0	335	2	G70369
13	210	16.9	337	2	A10079
14	209	16.9	324	2	F75029
15	209	16.9	337	2	C91122
16	209	16.9	337	2	B85967
17	206	16.6	337	1	Q0ECR6
18	205.5	16.6	348	2	D82807
19	204.5	16.5	534	2	H69056
20	202	16.3	324	2	C71215
21	201.5	16.2	421	2	T18825
22	201	16.2	354	2	C81040
23	200	16.1	325	2	A38108
24	199	16.0	354	2	C81986
25	195.5	15.8	340	2	B97011
26	195.5	15.8	346	2	F59766
27	195	15.7	353	2	T04567
28	195	15.7	412	2	T40899
29	193.5	15.6	327	2	G72411

30	193	15.6	348	2	S75548	staloglycoproteina
31	192	15.5	346	2	AF1820	staloglycoproteina
32	192	15.5	346	2	H70195	staloglycoproteina
33	189	15.2	336	2	C97888	O-sialoglycoprotei
34	188.5	15.2	344	2	AC1334	glycoprotein endop
35	186.5	15.0	344	2	H70737	probable O-sialogl
36	186	15.0	336	2	B95015	glycoproteinase fa
37	183	14.8	343	2	D83718	glycoprotein endop
38	182	14.7	338	2	A71545	probable O-sialogl
39	180.5	14.6	539	2	A64441	O-sialoglycoprotei
40	180	14.5	246	2	G90187	O-sialoglycoprotei
41	179	14.4	344	2	AB1705	glycoprotein endop
42	178	14.4	346	2	B84936	O-sialoglycoprotei
43	174.5	14.1	344	2	H72106	O-sialoglycoprotei
44	174.5	14.1	344	2	B86515	O-sialoglycoprotei
45	174.5	14.1	374	2	T35581	probable O-sialogl

ALIGNMENTS

RESULT 1
E84888
probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Jun-2003
C:Accession: E84888
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.;
M.; Xoo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MID:20083487; PMID:10617197
A:Accession: E84888
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1463 <STO>
A:Cross-references: GB:AF002093; NID:g2583127; PIDN:AAB82636.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g45270
A:Map position: 2
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 26.3%; Score 326; DB 2; Length 463;
Best Local Similarity 33.6%; Pred. No. 4,3e-22;
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 1 LIALVQGVDFLLGKSLDIAPQMDKVARRLSLIKHPGCSMGSKAIEHLAKQGNRF 60
221 LVLAKLQGYQLGTTVDALGEAFDTAKMGLDMH-----RSGPVAEELALEGDAK 275
DB 61 HEDIKRPLHARNKCDSPFGLQHTVDKIIIMKEKEBIEKGQILSSAADIATVQHTMAC 120
QY 121 HLYKRTTRAILPCKQKDLIPONNAVLVAGGVASNYITRALEITLNAOCTILCPPRL 180
DB 326 HHEKCKERAIDWALE---LEPSIKHNVISGVASNKYVRLNINVENKRLKALVCPSPS 382
QY 181 CTDNGIMTAMNGIERLRACILGILHDEGIRYE-----PRCPGLGVISKE 224
DB 383 CTDNGVMAWMTGLEHFRVG-----RYDPPPTPEBDYVDLRLPWPGLGEYSYAKG 432
QY 225 VGEA 228
DB 433 RSEA 436

RESULT 2
AB2902
O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AB2902
 R/Wood, D.W.; Serubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Moo, I.
 erage, G.; Gillet, M.; Grant, C.; Guentherer, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, F.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, W.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, B.W.
 A>Title: The genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AB2902
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-365 <KIR>
 A/Cross-references: GB:AE006688; PIDN:AA143632.1; PID:g17741154; GSPDB:GN00186
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A:Gene: gcp
 A:Map position: circular chromosome
 C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 365;
 Best Local Similarity 31.9%; Pred. No. 1.1e-14;
 Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

QY 2 LALVQGVSDPFLLGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIENHLAKQGNRFH 61
 DB 149 LVLVAGVGEYRWGTTTIDALGEAFDKTAKLGL-PYP-----GSPAVENNAAGDDPR 201
 62 FDIKPEPLHAKNCDPSFTGL-----QHVTDKIIMKKEKEGIEKGQILSSADIAATVQHT 117
 DB 202 FPLRPVWGEARLDPSFSGLKTAIVQAATATAPLSEQD-----INDICASFQA 250
 QY 118 MACHLVKTRHRAILFCRKORDLPONNA--VLVAGSVASNFYIRALBITVNAOTCTLIC 175
 DB 251 VSRLTKRIGRGLARFKVE--FPHINGEPALVAGVAVANQRIOTLQALCTHGFRAVA 308
 QY 176 PPRPLCTDNGIMIANNGIERLRAGLILHDIGIRYRCPGLGVDSKE 224
 DB 309 PPHRLCTDNAAIMIANWAGLERMABEG---RQADALBFAVPSRWPLODSAB 353

RESULT 3
 D97677
 Probable o-sialoglycoprotein endopeptidase (glycoproteinase) [imported] - *Agrobacterium*
 C/Species: *Agrobacterium tumefaciens*
 C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C/Accession: D97677
 R/Gooder, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*
 A/Reference number: A97359; MUID:21608551; PMID:11743194
 A/Accession: D97677
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-366 <KIR>
 A/Cross-references: GB:AE007869; PIDN:AAK88373.1; PID:g15157858; GSPDB:GN00169
 C/Genetics:
 A:Gene: AGR C 4806
 A:Map position: circular chromosome
 C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 366;
 Best Local Similarity 31.9%; Pred. No. 1.1e-14;
 Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

QY 2 LALVQGVSDPFLLGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIENHLAKQGNRFH 61
 DB 150 LVLVAGVGEYRWGTTTIDALGEAFDKTAKLGL-PYP-----GSPAVENNAAGDDPR 202
 62 FDIKPEPLHAKNCDPSFTGL-----QHVTDKIIMKKEKEGIEKGQILSSADIAATVQHT 117
 DB 203 FPLRPVWGEARLDPSFSGLKTAIVQAATATAPLSEQD-----INDICASFQA 251

QY 118 MACHLVKTRHRAILFCRKORDLPONNA--VLVAGSVASNFYIRALBITVNAOTCTLIC 175
 DB 252 VSRLTKRIGRGLARFKVE--FPHINGEPALVAGVAVANQRIOTLQALCTHGFRAVA 309
 QY 176 PPRPLCTDNGIMIANNGIERLRAGLILHDIGIRYRCPGLGVDSKE 224
 DB 310 PPHRLCTDNAAIMIANWAGLERMABEG---RQADALBFAVPSRWPLODSAB 354

RESULT 4
 AB3274
 O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - *Brucella melitensis* (strain
 C/Species: *Brucella melitensis*
 C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C/Accession: AB3274
 R/Delvecchio, V.G.; Kapatali, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Ios, T.; Ivanova, J
 .; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A/Reference number: AB3252; PMID:11756688
 A/Accession: AB3274
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <KIR>
 A/Cross-references: GB:AB008917; PIDN:AA151357.1; PID:g17982056; GSPDB:GN00190
 A/Experimental source: strain 16M
 C/Genetics:
 A:Gene: BME10175
 A:Map position: 1
 C/Superfamily: O-sialoglycoprotein endopeptidase
 C/Keywords: hydrolase; metalloproteinase

Query Match 19.5%; Score 241.5; DB 2; Length 359;
 Best Local Similarity 34.5%; Pred. No. 1.9e-14;
 Matches 69; Conservative 27; Mismatches 77; Indels 27; Gaps 6;

QY 2 LALVQGVSDPFLLGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIENHLAKQGNRFH 61
 DB 145 LVLVAGVGEYRWGTTTIDALGEAFDKTAKLGL-PYP-----GSPAVENNAAGDDPR 201
 62 FDIKPEPLHAKNCDPSFTGL-----QHVTDKIIMKKEKEGIEKGQILSSADIAATVQHT 116
 DB 198 FPLRPVWGEARLDPSFSGLKTAIVQAATATAPLSEQD-----INDICASFQA 245
 QY 117 MACHLVKTRHRAILFCRKORDLPQ--NNAVLVAGSVASNFYIRALBITVNAOTCTLIC 175
 DB 246 AVADLTSLDRVGSLSRFKTE--FPDCATPSLVVAGVAVANQRIOTLQALCTHGFRAVA 303
 QY 176 PPRPLCTDNGIMIANNGIER 195
 DB 304 PPHRLCTDNAAIMIANWAGLER 323

RESULT 5
 E71711
 Probable o-sialoglycoprotein endopeptidase (gcp) RP037 - *Rickettsia prowazekii*
 C/Species: *Rickettsia prowazekii*
 C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
 C/Accession: E71711
 R/Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ronten, T.; Alsmark, U.
 Nature 396, 133-140, 1998
 A>Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.
 A/Reference number: A71630; MUID:99039499; PMID:9823893
 A/Accession: E71711
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-387 <AND>
 A/Cross-references: GB:AJ235270; GB:AJ235269; NID:93860572; PIDN:CAA1450
 A/Experimental source: strain Madrid B
 C/Genetics:
 A:Gene: gcp; RP037
 C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 238; DB 2; Length 387;
Best Local Similarity 29.7%; Pred. No. 4.4e-14;
Matches 70; Conservative 32; Mismatches 74; Indels 60; Gaps 6;

QY 5 VQGVSDPLLKSLDIPGMDLKVARRSLIKHPECSMTSGSKAIEHLAKQGNRFHDI 64
DB VLGIKGRKILGTTTDDVAGEFDDKAKMLN-----SFGGPEIEKRAKIGNPKYKF 197
QY 65 KPLIHHAKNCDPSFTGLQHTVDKIIMK-KEKEGIEKGQILSSADIAATVQHTMACILV 123
DB 198 PKPIINSNCNMSFGLKTAVRTLIMLKEVNDV-----INDIASFPTGAILIS 249
QY 124 KRTTRAILFCQK-----RD-----LPQN----- 142
DB 250 SKMGDAIRLVQILNDYEDINHPKLNLSFRDGFWMKPLECTTRKRYRHIIONSTRS 309
QY 143 ---NAVLVSGVSNFYIRALBITNATQCTLLCPPRLCTDNGMIAMNGIER 195
DB 310 NLNDITVIAGVANKYLQELISDCTRPYGRILAPMHLCTDMAAMIAVAGLER 365

RESULT 6

Peptidase M22 family protein [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C/Accession: F87257
B/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: F87257
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-367 <STO>
A/Cross-references: GB:AE005673; NID:913421168; PIDN:AAK22058.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC0071
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 237.5; DB 2; Length 367;
Best Local Similarity 34.3%; Pred. No. 4.6e-14;
Matches 73; Conservative 29; Mismatches 90; Indels 21; Gaps 6;

QY 5 VQGVSDPLLKSLDIPGMDLKVARRSLIKHPECSMTSGSKAIEHLAKQGNRFHDI 64
DB 156 VSGVGACRKLGTITDDAGAEAFDKIAKSLG-PYP-----GPRLEKLVAGSDPTRYAL 208
QY 65 KPLIHHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQILSSADIAATVQHTMACILV 124
DB 209 PRALIGRKDCDFSPGSKTAARIAETLTTDD-----ARBDLAGVQAAIARQISE 259
QY 125 KTRTRAILFCQKRDLPQNNAVLVASGVASNFYIRALBITNATQCTLLCPPRLCTDN 184
DB 260 RVDRAKMLYK--DSHDEPDLAFVAVAGVAAVRAALLADCEKRGSPAPAPLAYCTDN 317
QY 185 GIMIANGIERLRAGLGLHDIEGIRYEPKPL 217
DB 318 AAMIALAGARL--ALGIFFDLDAIA-RPRWPL 347

RESULT 7

H83572

O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: H83572
R/Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.; Adaman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latdige, K.; Lim, J.; Loty, S.; Olson, M.V.

Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: H83572
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-341 <STO>
A/Cross-references: GB:AE004494; GB:AE004091; NID:99946446; PIDN:AAQ03969.1; GSPDB:GN001;
A/Experimental source: strain PA01
C/Genetics:
A/Gene: gcp
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 18.6%; Score 231; DB 2; Length 341;
Best Local Similarity 33.8%; Pred. No. 1.6e-13;
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

QY 2 LALVQGVSDPLLKSLDIPGMDLKVARRSLIKHPECSMTSGSKAIEHLAKQGNRFH 61
DB 142 LVRAVDIGRYQLLGESVDDAAGEAFDKTAKLIGL-GYP-----GGPEIARLAEGRTPGR 194
QY 62 FDIKPLIHHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQILSSADIAATVQHTMAC 121
DB 195 FVFPFPMTDPRGLDPSFGSLKTFITLN-TWQRCVEAGDDEQ--TRCDIALAFQTAVER 250
QY 122 LVKRTTRAILFCQKRDLPQNNAVLVASGVASNFYIRALBITNATQCTLLCPPRLC 181
DB 251 LLIKCRRL--KQGL--KN--LVLAGSVSNQALRSGLEKMGKQVYARPRFC 302
QY 182 TUNGIMIANGIERLRAGLGLHDIEGIRYEPKPL 217
DB 303 TUNGAMIANGICQRLLAG---QHDEPAISVQRRWPM 335

RESULT 8

E97707

O-sialoglycoprotein endopeptidase (BC 3.4.24.57) [imported] - Rickettsia conorii (strain E97707
C/Species: Rickettsia conorii
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C/Accession: E97707
R/Ogata, H.; Audic, S.; Renezo-Audiffren, P.; Fournier, P.B.; Barbe, V.; Samson, D.; Rol Science 293, 2093-2098, 2001
A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A/Reference number: A97700; MUID:21442074; PMID:11557893
A/Accession: E97707
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-344 <KUR>
A/Cross-references: GB:AB006914; PIDN:AAI02599.1; PID:915619097; GSPDB:GN00173
C/Genetics:
A/Gene: gcp
C/Superfamily: O-sialoglycoprotein endopeptidase
C/Keywords: hydrolase, metalloproteinase

Query Match 18.6%; Score 230.5; DB 2; Length 344;
Best Local Similarity 33.2%; Pred. No. 1.9e-13;
Matches 64; Conservative 34; Mismatches 76; Indels 19; Gaps 5;

QY 5 VQGVSDPLLKSLDIPGMDLKVARRSLIKHPECSMTSGSKAIEHLAKQGNRFHDI 64
DB 145 VLGIKGRKILGTTTDDVAGEAFDKVAKMLN-----AFPGGPEIEKRAKIGDPKRYKF 197
QY 65 KPLIHHAKNCDPSFTGLQHTVDKIIMK-KEKEGIEKGQILSSADIAATVQHTMACILV 123
DB 198 PKPIINSNCNMSFGLKTAVRTLIMLKEINDV-----INDIASFPTGAILIS 249
QY 124 KTRTRAILFCQK-RLLPQNNAVLVASGVASNFYIRALBITNATQCTLLCPPRLCT 182
DB 250 SKVQDAIRLVQILNDYEDINHPKLNLSFRDGFWMKPLECTTRKRYRHIIONSTRS 307
QY 183 DNGIMIANGIER 195
DB 308 DNAAAMIAVAGLER 320

RESULT 9

H64074

O:siatologlycoprotein endopeptidase (EC 3.4.24.57) - Haemophilus influenzae (strain Rd KW2)

N:Alternate names: siatologlycoproteinase

C:Species: Haemophilus influenzae

C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004

C/Accession: H64074

R:Reichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A/Accession: H64074

A>Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A:Residues: 1-342 <TIGR>

A:Cross-references: UNIPROT:P43764; GB:U32735; GB:I42023; NID:g1573509; PIDN:AAC22187.1;

C:Superfamily: O-siatologlycoprotein endopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match

Best Local Similarity 18.1%; Score 224; DB 2; Length 342;

Matches 67; Conservative 28; Mismatches 83; Indels 22; Gaps 6;

QY 2 LALVGVSDFLILGKSLDIAFGMDLVKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRFH 61

DB 142 LVRVDGVGVKYEIVGSIIDDAAGEAFDPTAKLLGL-DYP-----GGALLSLRAEKGTPRR 194

DB 62 FDIKPLHAKNCDPSFTGLQHVTDKIIKKKEKEG--IEKQIISSADIAATVQHTMA 119

DB 195 FTFRPMTDRAGLPDSFSGLTFTFPAANTVNOAIKNEGELIEQ-----TRADIAVAFQDAV 249

QY 120 CHLVKRTTRAILFCKQRDLFPONNAVIVASGVASNFYIRRALITLTNAQTCLCPPR 179

DB 250 DTLA-----IKCK-RALKETGKRLVIVAGSVANSKKLRETLAHLMLNGEVEVYPPQ 301

DB 180 LCTDNGIMIANNGIERLRAG 199

DB 302 FCTDNGAMIAVTFGLRLKQ 321

RESULT 10

G69388

O:siatologlycoprotein endopeptidase homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000

C/Accession: G69388

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utermbeck, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: G69388

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A:Residues: 1-323 <KMB>

A:Cross-references: GB:AE001027; GB:AE000782; NID:g2689350; PIDN:AA890129.1; PID:g264947

C:Superfamily: O-siatologlycoprotein endopeptidase

Query Match

Best Local Similarity 17.9%; Score 221.5; DB 2; Length 323;

Matches 62; Conservative 30; Mismatches 66; Indels 31; Gaps 8;

QY 11 FLILGKSLDIAFGMDLVKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIPLH 70

DB 142 YRVGRTLDIGIALMDKLARHML-KNP-----GPKRIEELAKKGQKTH--LP--YV 190

QY 71 AKNCDPSFTGLQHVTDKIIKKKEKEGKQILSSAADIATVQHTMACHLVKTRTAL 130

DB 191 VKGMDPSFSGMTAAQRLE-----DSGVA-----MEVVASFOSTAPAMLTETVERL 238

QY 131 LFCQRDLFPONNAVIVASGVASNFYIRRALITLTNAQTCLCPPRICDNGIMIAM 190

DB 239 AV-----LDLAEVLLV--GGVAANRLQEMLRIMEDRGAKFYVPEKELAGDNGAMIAV 290

QY 191 NCIEERLRAG 199

DB 291 TGLIMWKHG 299

RESULT 11

AG0892

probable glycoproteinase [imported] - Salmonella enterica subsp. enterica serovar Typhi (st

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: AG0892

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AG0892

A>Status: preliminary

A/Molecule type: DNA

A:Residues: 1-337 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD07733.1; PID:g16504285; GSPDB:GN00176

C:Genetics:

A:Gene: STY3387

C:Superfamily: O-siatologlycoprotein endopeptidase

Query Match 17.2%; Score 213; DB 2; Length 337;

Best Local Similarity 30.6%; Pred. No. 7,4e-12;

Matches 66; Conservative 32; Mismatches 92; Indels 26; Gaps 7;

QY 2 LALVGVSDFLILGKSLDIAFGMDLVKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRFH 61

DB 142 LLSVIGIGYELGSIIDDAAGEAFDPTAKLLGL-DYP-----GGPMLSKVASQSTAGR 194

QY 62 FDIKPLHAKNCDPSFTGLQHVTDKIIKKKEKEGKQILSSAADIATVQHTMACH 121

DB 195 FVPRPMTDRPGLDPSFSGLTFTFPAANTIRNSGDDE-----QTRADIAVAFEDAVDT 246

QY 122 LVKRTTRAILFCKQRDLFPONNAVIVASGVASNFYIRRALITLTNAQTCLCPPRIC 181

DB 247 L-----MIKKK-RALBSTGFKRLVMAGVANSRLTRAKLLEMMQKRGEVYFARPERC 298

QY 162 TDNGIMIANNGIERLRAGIILHDIEGIRYERKCP 217

DB 299 TDNGAMIAVAGNVRFKA--GVTADL-GVTVRPRWLP 331

RESULT 12

G70369

O:siatologlycoproteinase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004

C/Accession: G70369

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oye

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: G70369

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A:Residues: 1-335 <AQF>

A:Cross-references: UNIPROT:O66986; GB:AB000708; NID:g2983356; PIN:AA06951.1; PID:g2298
A:Experimental source: strain VFS
A:Genetics: gcp
A:Gene: gcp
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 17.0%; Score 210.5; DB 2; Length 335;
Best Local Similarity 32.0%; Pred. No. 1.2e-11;
Matches 70; Conservative 41; Mismatches 73; Indels 35; Gaps 10;

OY 2 LALVGVSDFLLEKSLDIAPGDMLDKVARRLSLIKHPECSMTSGSKAIEHLAKOGRFH 61
DB 141 LYLVRDGRVDFLGGLTDDAAGEAYDKVAKMLGL-GYP-----GGPIIDLAKKEKKL- 192

OY 62 FDIKPLPAAKNCDFSTGLQHTVDKLIIMKKEKEGIEKGQILSSADIAATVQHTMACH 121
DB 193 YPLPKPLMEGNNINFSFSGLK---TAIINLTKKEKNVRK-----EDIAVSFOETVVEI 242

OY 122 LVKTRTRAILFCQKORDLLPQNNAVLVASGVASNYIRALAILTNATQ---CTLLCPPL 178
DB 243 LLEKSK---LWAMKKTGIKR---LVVGVGSANRLR---EVFKKASQEGFELYIPIHP 291

OY 179 RLCTDNGIMIAWNGIERLRAGILGILHIDIEGIRYEPKCPPL 217
DB 292 SLTDNAMLTAIYAGMERFKGVAPLDVNP---QINIRL 327

RESULT 13
AI0079
probable glycoproteinase gcp [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence, revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AI0079
G:Parhill, J.; Wren, B.W.; Thomson, N.R.; Tilball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Ellis, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
NATURE 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AI0079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KIR>
A:Cross-references: GB:AL590842; PIDN:CAC89500.1; PID:g15978736; GSPDB:GN00175
C:Genetics: gcp
A:Gene: gcp
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 210; DB 2; Length 337;
Best Local Similarity 30.0%; Pred. No. 1.4e-11;
Matches 65; Conservative 33; Mismatches 91; Indels 28; Gaps 6;

OY 2 LALVGVSDFLLEKSLDIAPGDMLDKVARRLSLIKHPECSMTSGSKAIEHLAKOGRFH 61
DB 142 LLSVTGIGELTIGESVDDAAGEAPDKTAKLGL-DYP-----GSPMLSRMAQQTGVR 194

OY 62 FDIKPLPAAKNCDFSTGLQHTVDKLIIMKKEKEGIEKGQILSSADIAATVQHTMACH 121
DB 195 FTRPRPTMDRGDIFSFGKLTFAANTIRANGDD-----QTRADIAAFEDAVVPT 246

DB 122 LVKTRTRAILFCQKORDLLPQNNAVLVASGVASNYIRALAILTNATQCTLLCPPL 180
247 LAIKSKRA-----LDQGFRLVYIAGGVASANQTLRLKLADMMQKRGGEVYARPEF 297

OY 181 CTDNGIMIAWNGIERLRAGILGILHIDIEGIRYEPKCPPL 217
DB 298 CTDNGAMITAYAGMVRKSN---INSELSSVSVPRPPL 331

RESULT 14
P75029
O-sialoglycoprotein endopeptidase (gcp) PAB1159 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi

C:\Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

R:\Accession: F75029
R:\Anonymouse, Genoscope
submitted to the EMBL Data Library, July 1999
A:\Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A:\Reference number: A75001
A:\Accession: F75029
A:\Status: preliminary
A:\Molecule type: DNA
A:\Residue: 1-324 <KAW>
A:\Cross-references: UNIPROT:Q9UXT7; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB5067(6)
A:\Experimental source: strain Orasay
C:\Genetics:
A:\Gene: gcp, PAB159
C:\Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 324;
Best Local Similarity 32.5%; Pred. No. 1,6e-11;
Matches 65; Conservative 34; Mismatches 69; Indels 32; Gaps 8;

OY 1 LIALVGVSDFLILGSLDIAPGDMLDKVARRLSLIKHPESCTWSGGKAIEHLAKOGNRF 60
 :||| : ||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 135 VVALLEG--RYRVFGETLDIGIGNAIDVFAREIQL-----GPGGPVKYEKLAEKEKY 185

OY 61 HDIRPLHAAKNCDPSFTGLQHVTDKIMKKKEBEGIEKGQILLSSADIAATVQHTMAC 120
 :-:-:-IELP-AAVGGMDSFSGL-LTEAIRKRRSGKVVE-----DLAVSQETAFA 230
DB 186 ---ELP-AAVGGMDSFSGL-LTEAIRKRRSGKVVE-----DLAVSQETAFA 230

OY 121 HLVRTHRALIFCKQRDLPONNAVVLASGVASNFIYRALLETITNATQTLLCPPRLL 180
 ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 231 ALVERTERVATE-----KDENVLV--GVVAANRLRELMILMTEDRGKPFVPYYDL 282

OY 181 CTDNGIMIAMNGIERLAGL 200
 || || || || || || || || || || || || || || || || || || || || ||
DB 283 CRDNQMAMAYTGRLMYKAGI 302

RESULT 15
C91122
probable O-sialoglycoprotein endopeptidase [imported] - Escherichia coli (strain O157:H7,
C:\Species: Escherichia coli
C:\Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:\Accession: C91122
R:\Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; H.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hatoczi, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:\Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:\Reference number: A99629; PMID:21156231; PMID:11258796
A:\Accession: C91122
A:\Status: preliminary
A:\Molecule type: DNA
A:\Residue: 1-337 <HAY>
A:\Cross-references: GB:BA000007; PIDN:BAB37370.1; PID:q13363420; GSPEB:GN00154
A:\Experimental source: strain O157:H7, substrain RMD 0509952
C:\Genetics:
A:\Gene: BC83947
C:\Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 337;
Best Local Similarity 30.6%; Pred. No. 1.7e-11;
Matches 66; Conservative 34; Mismatches 90; Indels 26; Gaps 8;

OY 2 LALVGVSDFLILGSLDIAPGDMLDKVARRLSLIKHPESCTWSGGKAIEHLAKOGNRF 61
 ||| : ||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 142 LISVTCIGCYELLGSIDDAAGEARPKTKLGL-DYP-----GGPLSKMAAQTAGR 194

OY 62 PDIKPLHAAKNCDPSFTGLQHVTDKIMKKKEBEGIEKGQILLSSADIAATVQHTMAC 121
 ||| : ||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 185 FYRPDMTRPEIDNEPCSTLTFFANTTI---TNQIQD---QIRADIARAEEADVUT 246

OY 122 LVKRTHRALIFCKQRDLPONNAVVLASGVASNFIYRALLETITNATQTLLCPPRLL 181
 ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 247 LMICKRRAL----DLNGFKR--LVMAAGVSANRTLRALGAEMMKGRGEVVYARBEFC 298

```

QY      182 TDNGIMAMNGIERLRAGGCI LHDIEGIRYEPKCP 217
          ||| ||| : : | : : ||
Db      299 TDNGAMIAVAGMRFKA--GATADL-GVSVPRMPL 331

```

Search completed: February 16, 2005, 13:09:23
Job time : 11.9109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2005, 13:19:07 ; Search time 2839.68 Seconds
(without alignments)
3303.653 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 L1ALVGVSDPFLIKGSLDI.....DISKVGESAIKVPOLKMEI 239

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 6847908

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-Q/cgnt2_1/USPTO.spool/US10649273/runat_14022005_114703_16411/app_query.fasta_1.1429
-DB=EST -QPM=fastcap -SUFFIX=xtc -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPM=pro -NOR=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273.QCGN_1.1_6799_@runat_14022005_114703_16411 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1118	90.2	870	5	BQ423651 AGENCOURT
2	1097	88.5	640	5	BQ636028
3	1079	87.0	2284	3	AK045669 Mus muscu
4	1071	86.4	1622	3	AK011265 Mus muscu
5	930	75.0	852	5	BX391919 BX391919
6	925	74.7	1603	2	BC030671 Mus muscu
7	908	73.2	701	2	BE740611 601595739
8	876	70.6	658	7	CP362328 829596 MA
9	838	67.6	637	7	CK941819 4065407 B

10	799	64.4	922	5	BQ961028	BQ961028 AGENCOURT
11	789	63.6	545	1	AV602901	AV602901 AV602901
12	784	63.2	490	6	CB852881	CB852881 UT-CF-FNO
13	783	63.1	822	7	CF257246	CF257246 Bha008_50
14	753	60.7	866	5	BU127463	BU127463 603114407
15	749	60.4	736	7	CK365185	CK365185 AGENCOURT
16	746	60.2	723	5	BU261281	BU261281 603502215
17	731.5	59.0	792	5	BU403563	BU403563 604138456
18	717	57.8	634	2	AM601179	AM601179 RC1-BT025
19	708	57.1	736	6	CA057753	CA057753 esal-rcb54
20	707.5	57.1	701	5	BU621780	BU621780 UT-H-FL1-
21	698	56.3	696	5	BU302606	BU302606 603739448
22	685	55.2	597	6	CB272391	CB272391 ma157912
23	683	55.1	909	5	BX756548	BX756548 BX756548
24	682	55.0	548	7	CO880741	CO880741 Bvgan_09
25	671	54.1	682	2	BM0433703	BM0433703 BE043703
26	671	54.1	730	7	CN823245	CN823245 Oa-ep1bn
27	668	53.9	706	5	BU202465	BU202465 603949052
28	666	53.7	869	5	BX754527	BX754527 BX754527
29	647	52.2	919	7	CF407294	CF407294 CH3#047_E
30	646	52.1	484	1	AJ670918	AJ670918 AJ670918
31	646	52.1	878	5	BX776940	BX776940 BX776940
32	642	51.8	1173	6	CD508917	CD508917 CDA93-E05
33	638	51.5	1082	5	BX359023	BX359023 BX359023
34	625	50.4	506	2	BF415802	BF415802 UT-R-CA1-
35	614	49.5	861	5	BU246158	BU246158 603779906
36	611	48.3	1171	5	BU261605	BU261605 603501763
37	604	48.7	424	1	AA273921	AA273921 v099c03_7
38	600	48.4	863	5	BU376295	BU376295 603808890
39	594	47.9	389	4	BM744822	BM744822 K-EST0018
40	591	47.7	812	5	BU246489	BU246489 603784202
41	589	47.5	357	5	BO672554	BO672554 AGENCOURT
42	581	46.9	704	5	BU327284	BU327284 603491570
43	576	46.5	826	7	CR444994	CR444994 CR444994
44	573	46.2	385	1	AJ647827	AJ647827 AJ647827
45	572.5	46.2	613	1	AA920105	AA920105 v25h05_x

ALIGNMENTS

RESULT 1
BQ423651 870 bp mRNA linear EST 23-MAY-2002
LOCUS AGENCOURT_7790948 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065828
DEFINITION BQ423651.1 GI:21118966
5', mRNA sequence.

ACCESSION BQ423651
VERSION BQ423651.1 GI:21118966
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 870)

AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCPD/BTP

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL3342 row: 1 column: 21
High quality sequence stop: 710.
Location/Qualifiers

FEATURES

source
1..870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6065828"

/tissue type="melanotic melanoma"
/lab host="DH10B (phage-resistant)"
/clone lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SORT6; Site_1: NotI,
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN

Alignment Scores:

Pred. No.: 3e-119 Length: 870
Score: 1118.00 Matches: 220
Percent Similarity: 98.22% Conservative: 1
Best Local Similarity: 97.78% Mismatches: 2
Query Match: 90.16% Indels: 2
DB: 5 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BQ423651 (1-870)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIle 20
DB 196 CTGTTGGCATTAGTTCAAGAGTTTCAATTTCTGCTTCTTGAAAGCTTTGGACATA 255
QY 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 256 GCACCAAGTGACATGCTTGGACAAGTGCCAGAGAACCTTCTTATATAAACATCCAGAG 315
QY 41 CysSerThreMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 316 TGCCTCCACCATGAGTGCGGAGAAAGCCATAGAACCTTTGGCCAAACAGAAATAGATT 375
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheArg 80
DB 376 CATTTTGACATCAAACTCCCTTGATCATGCTAAATAATGATTTTCTTTACTGGA 435
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlyGluGlyIleGluLys 100
DB 436 CTTCAACACGTTATGATATAATAATATGAAAAAGAAAAAGAGAGTTTGAAGAAG 495
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
DB 496 GGGCAATTCCTGCTTCCAGACAGACATGCTGCCACAGTACACACACATGGCATGT 555
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysGlyGlnArgAspLeuLeuPro 140
DB 556 CATCTTGTGAAAAACACATCGGCTATTCTGTTTGAACAGACAGACATGTTACT 615
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
DB 616 CAAAATATATGACATGATGCTGTCATCTGTCGTGTCGACAGTAACTTCTATATCCGACGA 675
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 676 GCTCTGGAATTTTAAACAAACGACACACATGTCCTTGGTGTCTCTCCGACATA 735
QY 181 CysThrAspAsnGlyIleMetIleAlaTTPAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 736 TGCATATATATGACATTAATGATGATGCAATGATGTAATGAAAGCTACGTCTGCTTGG 795
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrglu-ProLysCysPro-LeuGlyValA 220
DB 796 GGCATTTTACATGACATGAAAGGCAATCGCTATGAAACCAATGGTCTCTTGGAGTAG 855
QY 220 spileserLys 223
DB 856 ACATATCAAAA 866
RESULT 2
BQ636028
LOCUS BQ636028 640 bp mRNA linear EST 15-JUL-2002
DEFINITION h003d11.y1 Human Retina cDNA (Un-normalized, unambified): hd/he
Homo sapiens cDNA clone h003d11 5', mRNA sequence.
ACCESSION BQ636028

VERSION

BQ636028.1 GI:21760487
EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 640)
Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A.,
Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.

AUTHORS

Expressed sequence tag analysis of human retina for the NRI Bank
Project: Retbindin, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts

TITLES

Mol. Vis. 8 (4), 196-204 (2002)

JOURNAL

22103461

MEDLINE

12107411

PUBMED

Contact: Wistow G

COMMENT

Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: gwaem@helix.nih.gov
Plate: 03 row: d column: 11
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers

FEATURES

source

1..640
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hd03d11"
/dev_stage="Adult"
/tissue_type="Retina"
/lab_host="EMDH10B"
/clone_lib="Human Retina cDNA (Un-normalized,
unambified): hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
was dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was constructed at
Bioserve Biotechnology (Laurel MD) essentially following
the protocols of the Superscript Plasmid System full
details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor
[5'-GGACTAGTTCATGATCGGAGCGGCCGCTT)15-3']. EST analysis
was performed on the unambified library at the NIH
Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.: 5.41e-117 Length: 640
Score: 1097.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.47% Indels: 0
DB: 5 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BQ636028 (1-640)

QY 31 ArgArgLeuSerLeuIleLysHisProGluCysSerThreMetSerGlyGlyValAlaIle 50
DB 3 AGAAGACCTTCTTATATAAACATCCAGAGTGCTCCACATGATGATGGGAAAGCCATA 62
QY 51 GlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHis 70
DB 63 GAACATTGGCCAAACAGAAATATGATTTTCAATTTGACATCAACCTCCCTGCATCAT 122
QY 71 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 90
DB 123 GCTAAAAATATGATTTTCTTTTACTGACACTTCAACAGCTTACTGATATAATATATATG 182

QY 91 Lysylsglytylsylglutylglytylglutylglytylglutylserinylalanylalanyl 110
 Db 183 AAAAAGAGAAAAAGAGAAAGATGAGAAAGGGGAAATCGTCTTCAGACAGACAACTT 242
 QY 111 Alalathrvalaglnhisthrmetacetyhistleuvalylargthrhialalile 130
 Db 243 GCTGCCACAGTACAGACACAAATGGATGTCATCTTGAAAAAGAACATCGGGCTATT 302
 QY 131 Leuphycysylsylvglatrgaspleuueuproglnaenanaalavalienvalalasergly 150
 Db 303 CTGTTTGTGTAAGCAGAGAGACTTGTACTCTAAATAATGCAATCTGGTTCATCTGGT 362
 QY 151 Glyvalalaserasphenetyrilleargrghalaleuglnileuethrasnalathrgln 170
 Db 363 GGTCCTCCAAATGACTTCTATATCCGACAGCTCTGAAATTTTAAACAGCAACAG 422
 QY 171 Cyethrleuueucyeproppropargleucyethrarpapenglyllemetlealatr 190
 Db 423 TGCACTTGTGTGTCTCCCTCCAGACTATGCACTGATATGGCATTATGATTCATGG 482
 QY 191 Asnglylilegluaargleuargalaglyleuuglylileuhsasppilegluglylilearg 210
 Db 483 AATGGTTTGAAGAACTACGCTGCTGGCTTGGCATTTTACATGACATGAGAGCAATCCGC 542
 QY 211 TyrgluProlysCyBProleuglyValaspiIeserlysgluvalglylualaserile 230
 Db 543 TATGAACCAAAATGTCTCTTGGATGACATATCAAAAGAAAGTTGGAGAGCTTCATTA 602
 QY 231 LysValProgluLeuLysMetGluile 239
 Db 603 AAAGTACCACAAATTAATAATGAGATA 629

RESULT 3
 AK045669 2284 bp mRNA linear HTC 03-APR-2004
 LOCUS AK045669
 DEFINITION Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B230219017 product:similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo sapiens), full insert sequence.
 ACCESSION AK045669
 VERSION AK045669.1 GI:26337528
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL MEDLINE 20499374
 PUBMED 11042159

REFERENCE
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitabumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 JOURNAL MEDLINE 11811111
 PUBMED 11511111

REFERENCE
 5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 JOURNAL MEDLINE 12011111
 PUBMED 12011111

REFERENCE
 6 (bases 1 to 2284)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kashiwagi, T., Kato, H., Kawai, Y., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submision
 Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@resc.riken.jp, URL: http://genome.resc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.resc.riken.jp/
 URL: http://fantom.resc.riken.jp/
 Location/Qualifiers
 1..2284
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:B230219017"
 /db_xref="taxon:10090"
 /clone="B230219017"
 /sex="male"
 /tissue_type="corpora quadrigemina"
 /dev_stage="adult"
 /dev_stage="adult"
 /note="unnamed protein product; putative similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 [Homo sapiens] (SPTRLQ9H4B0, evidence: PASTY, 80%ID, 100%length, match=142)"
 /codon_start=1
 /protein_id="BAC32450.1"
 /db_xref="GI:26337529"
 /translation="MTMGRRGALPKPKSKVYGLRPRVHPRTISCHKLVIGIETSCDDGAIVDRTGVNGLGALHSQTVHRLKGTGIVPPVAAQVHRENIQIVERTLSACRITPDSIAIATITTEGLALSGVLSPLQVNVNPFKEPIIHMEALATITRLTNKVBEPVLVILSGCLLALVQVSPFLILGSLDIAPGMDLRVARSLIRHPESTMSGGAIAIOLADQGRFHTIPRNQNAKQSPFGLOHITDKLITHEKEGIEKGQILSGAADIILAAVQHTAHLAKRTHRALIPKQKNLSPANAIVVSGVASNIYIRKALERVANAQTCLCPPRCTDNGIMIANNGIERLRAGLVLDVEDIRERPCPLGVDSIRVAAEALIKVPRLEKML"
 /note="2267"
 /note="putative"
 2284
 /note="putative"

FEATURES
 source
 polyA_site
 polyA_signal
 ORIGIN
 11076861

Source	1..1622
organism="Mus musculus"	
mol_type="mRNA"	
strain="C57BL/6J"	
/db_xref="PANTOM DB:2610001M19"	
/db_xref="taxon:10090"	
/clone="2610001M19"	
/issue_type="whole body"	
/clone_id="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="10 days embryo"	
207..1451	
/note="unnamed protein product; putative	
similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo	
sapiens) (SPTR Q9H4B0, evidence: PASTY, 80%ID, 100%length,	
match=1242)"	
/codon_start=1	
/protein_id="BAB27506.1"	
/db_xref="GI:12847276"	
/translation="MLMRRPAGAIIPKPKSKVGYPLRFSVHPRTLSCHLVIGITST	
SCDDTGAAYVDENUGGEALHSOTQVHLTKGVIYPPAQDLHENTORIYEBTSLAC	
RITPSDLSAIRITTKPGGLATSLGVGLSFLDVQPKPFIPIHMEHALTTLNTRK	
VERPFLVLLISGGICLLALVGVSDPFLILGSLDIAIPDMIDLKARLSLTKPECSBT	
MSGGALIEOLAKDGNRHFITNPPMONAKNCDPFGLOIHTDKLITKEKEBIEKEE	
QISSAADIDAAVOHATACHTAKRTHRIELFCOKNLSPPANAVLVVSGGVAANLYLR	
KALEIVANQCTLLCEPRLCITDNGIMIANNGIERLRAGIVLHVEDIRYEKCPDL	
GVDISREYEAIRIVPRKAL"	
1605..1610	
/note="putative"	
1622	
/note="putative"	
/note="putative"	

OY 161 AlaleugluileuThrrAaAaIaThGlcCyethrleuLeuCySProProAryLeu 180
 Db 1212 GCATTGGAATTTGCGCAATGCAACGACGCTGTTGTGTCCACTCCCAAGACTG 1271
 OY 181 CyethrAaPaanglyIilemeTleAlaITrPaanglyIilegluaLeuAryAlaIglYleu 200
 Db 1272 TGCACTGACATGATGCATCATGATTCATGGAATTCGAATTCGAATTAACGTGCTGCTTG 1331
 OY 201 G1Y1leleuHsaApIilegIunglyIilaayTTCgluProlyCySProleuGlyValAsp 220
 Db 1332 GGCCTTTTACATGATGTGAGAAACATCCGATGACCAAAATATCTCTTGGAGTGAAC 1391
 OY 221 IleserlyegIuValIglYgluaIaSerIilelyeValProGInleuLySMeTgluIle 239
 Db 1392 ATATCCAGAGAAAGTTGCAGAAAGCTGCCATTAATAATACCGCATTTAAATAATGCACTT 1448
 RESULT 5
 LOCUS BX391919/c 852 bp mRNA linear EST 28-APR-2004
 DEFINITION BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 cDNA clone CS0DK001YE02 3-PRIME, mRNA sequence.
 ACCESSION BX391919
 VERSION BX391919.2 GI:46846154
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 852)
 L.J.W.B., Gruber,C., Jesse,J., and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On May 13, 2003 this sequence version replaced gi:30611736.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 Evry cedex - FRANCE
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 1240.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?cs=CS0BA1035Z07_CS03317_1kc=1240.r

FEATURES
 source
 1. 852 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK001YE02"
 /cell_type="HELA CELLS COT 25-NORMALIZED"
 /cell_line="HELA"
 /clone_1lb="Homo sapiens HELA CELLS COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ALIGNMENT SCORES:
 PRED. NO.: 2,28e-97 Length: 852
 Score: 930.00 Matches: 180
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 75.00% Indels: 0
 DB: 5 Caps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BX391919 (1-852)

```

Db      542 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGGAAAGCTTTGGACATA 483
Qy      21  ALaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db      482 GCACCAAGTGCATGCTTGACAGAGGTGCAGAGAACCTTTCTTATATAAACATCCAGAG 423
Qy      41  CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGluGlyAsnArgPhe 60
Db      422 TGCTCCACCATGATGAGTGTGGAAAGCCATGAAACCTTTGGCCAAACAGAAATGATTT 363
Qy      61  HisPheAspIleLysProProLeuHisIleAlaLysAsnCyAspPheSerPheThrGly 80
Db      362 CATTTCGATCAACAACCTCCCTGCATCATGCTAATAAATTTGATTTTCTTTACTGGA 303
Qy      81  LeuGlnHisValAlaThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
Db      302 CTTCAACACGCTTACTGATTAATAATATATGAAAAAGAAAAAGAGATTTGAGAA 243
Qy      101 GLYGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db      242 GGGCAATCTCTGCTTCCAGACAGACATGCTGCGACAGTACAGACACAAATGGCATGT 183
Qy      121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db      182 CATCTTGGAAGAAAGACACATCGGCTATTCTGTTTGTAAAGACAGAGACTGTTACT 123
Qy      141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAsnAsnPheTyrIleArgArg 160
Db      122 CAAATATATGACAGTACTGTTGATCTGTGTGTGTCGACAGTAACTTCTATATCCGAGA 63
Qy      161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db      62 GCTCTGAAATTTTAAACAACGACACAGATGCTACTGTTGTGTCTCTCCACAGACTA 3

```

```

RESULT 6
LOCUS   BC030671          1609 bp    mRNA    linear    HTC 19-NOV-2003
DEFINITION Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA
ACCESSION BC030671
VERSION   BC030671.1
KEYWORDS HTC.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stadleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schaefer, T.E., Brownstein, M.J., Umeda, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Kana, S.S., Lequellano, N.A., Peters, G.J.,
Adamson, R.D., Mullihy, S.J., Bosak, S.A., McEwan, P.J.,
McMernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Vallalath, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.B.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bonifield, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Gilmwood, J., Schmitz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL MEDLINE
PUBMED 12477932
REFERENCE 2 (bases 1 to 1609)
AUTHORS Strausberg, R.

```

TITLE
Direct Submission
JOURNAL
Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
CDNA Library Preparation: M. Bento Soares, University of Iowa
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadansystemsbiology.org
Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES
source
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAX Plate: 66 Row: e Column: 10
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463
This clone has the following problem: frame shifted.
Location/Qualifiers
1..1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_xref="IMAGE:1226118"
/tissue_type="Thymus gland, mouse"
/clone_lib="Soares_thymus_2NMNT"
/lab_host="DH10B"
/note="Vector: pT73-Pac"

ORIGIN

Alignment Scores:

```

pred. No.:      1 65e-96      length:      1609
Score:          926.00      Matches:      183
Percent Similarity:      81.59%      Conservative: 12
Best Local Similarity:    76.57%      Mismatches:   20
Query Match:      74.68%      Indels:       24
DB:              3          Gaps:          1

```

US-10-649-273-2_COPY_176_414 (1-239) x BC030671 (1-1609)

```

Qy      1  LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db      749 CTGTTGGCATTGATTCGAAGGTGTTCCGATTTCTCTCTCTGGAAAGCTTTGGACATA 808
Qy      21  ALaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db      809 GCACCAAGGCGACATGCTTGACAAAGTGCAGAGAACTTTCTTATATCAAAACATCCAGAA 868
Qy      41  CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGluGluGlyAsnArgPhe 60
Db      869 TGCTCCACCATGATGAGTGTGGAAAGCCATGAAACCTTTGGCCAAACAGAAATGATTT 928
Qy      61  HisPheAspIleLysProProLeuHisIleAlaLysAsnCyAspPheSerPheThrGly 80
Db      929 CATTTCGATCAACAACCTCCCTGCATCATGCTAATAAATTTGATTTTCTTTACTGGA 988
Qy      81  LeuGlnHisValAlaThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
Db      989 CTTCAACATATATGATGATTAATTAACAACAGAAAAAGAAAAAGAGCATTTGAGAA 1048
Qy      101 GLYGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db      1049 GGGCAATCTCTGCTTCCAGACAGACATGCTGCGTGTACAGATGACACAGAGCTGC 1108
Qy      121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db      1109 CACTTTCGAAAGAAAGACATCGGCTATTCTGTTTGGACAGAGAAAAATTTGCTCTCT 1168

```

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArgArg 160
 DB 1169 CAGCTAAACGAGATTAGTGTATCTGGAGGTGTGCAAGTAATTGACATCGGAAA 1228
 QY 161 AlaleuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgIleu 180
 DB 1229 GCATTGGAAATTTGCGCAAAATGCAACGACGATGCTGTTGTGTCACCTCAAGACTG 1288
 QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgIleu 200
 DB 1289 TGACCTACCAATGACATCATATATGCA----- 1315
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
 DB 1316 -----TGATGCTCTTGGATGAC 1336
 QY 221 ILeSerIysGluValAlaGlyValAlaSerIleLysValProGlnLeuLysMetGluIle 239
 DB 1337 ATATCCAGAGAGATTGCGAGAGCTGCCATTAAGTAAAGTAAATGCGCACTT 1393
 RESULT 7
 LOCUS BE740611 701 bp mRNA linear EST 15-SEP-2000
 DEFINITION 601595739P1 NIH_MGC_9 Homo sapiens CDNA clone IMAGE:3949640 5',
 mRNA sequence.
 BE740611
 ACCESSION BE740611.1 GI:10154603
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 701)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: gcapbs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LML at: Image.lml.gov
 Plate: L10614 row: n column: 09
 High quality sequence stop: 701.
 Location/Qualifiers
 1..701
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3949640"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_9"
 /note="Organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.31e-95 Length: 701
 Score: 908.00 Matches: 178
 Percent Similarity: 58.91% Conservative: 3
 Best Local Similarity: 97.27% Mismatches: 0
 Query Match: 73.23% Indels: 2
 DB: 2 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BE740611 (1-701)
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
 DB 155 CTGTGGCAATGATGTTCAAGAGATTTCAGATTTCTCTTGTGAAAGCTTTGGACATA 214
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysIleSPGlu 40
 DB 215 GACACGAGTGAATGCTTGAACAGTGGCAAGAGACTTTCTTAATTAACATCCAGAG 274
 QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
 DB 275 TGTCTCACCATTAGATGTGTGGAAAGCCATAGACATTGGCCAAACAGAAATATGTTT 334
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheTrgIy 80
 DB 335 CATTTTGACATCAAACTCCCTTGATCATGCTTAAAAATTTGATTTTCTTTACTGGA 394
 QY 81 LeuGlnHisValThrAspLysIleIleMetLys--LysGluLysGluGluGlyIleGlu 100
 DB 395 CTTCAACACGTTACTGATTAATTAATTAATGAACAGAAACAGAAAGTATTGAGA 454
 QY 100 YAGIYGLInileuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaC 120
 DB 455 AGGGGCAAAATCTGTCTTACGACGACGACATTCCTCCACAGTACACACAAATGGCAT 514
 QY 120 YHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuArg 140
 DB 515 GTCATCTGTGTAAGAAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTGTAC 574
 QY 140 TCGIAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArgArg 160
 DB 575 CTCAAAATTAATGACGATGCTGTCATCTGCTGTCGTCGTCGTCGTCGTCGTCGTCG 634
 QY 160 rGAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgIle 180
 DB 635 GAGCTCTGGAATTTTAACAAACGACACAGTGCCTTTGTTGTCTCTCTCCACAGAC 694
 QY 180 euCys 181
 DB 695 TATGC 699
 RESULT 8
 CF362328/c 658 bp mRNA linear EST 25-AUG-2003
 LOCUS 829596 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.
 DEFINITION CF362328
 ACCESSION CF362328.1 GI:34161882
 VERSION EST.
 KEYWORDS Sus scrofa (pig)
 SOURCE Sus scrofa
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 REFERENCE 1 (bases 1 to 658)
 AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
 Wise,T.A., Noneman,D.J., Wray,J.B. and Keele,J.W.
 TITLE A second set of porcine ESTs from a pooled-tissue normalized
 library
 JOURNAL Unpublished (2003)
 COMMENT Contact: Smith TP
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemail.marc.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_alt option. Vector identified with
 cross_match v0.990329.
 Plate: SRG8015 row: H column: 12
 Seq primer: TAGAAGCAGCATGTCAGC.
 Location/Qualifiers
 1..658
 FEATURES
 source

```

/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_11b="MARC 3P1G"
/notice=Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium
ovary, testes, and bone marrow."

```

ORIGIN

Alignment Scores:	
Pred. No.:	3, 07e-91
Score:	876.00
Percent Similarity:	93.68%
Best Local Similarity:	86.84%
Query Match:	70.65%
DB:	7
Length:	656
Matches:	165
Conservative:	13
Mismatches:	12
Indels:	0
Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) X CF362328 (1-658)

Qy	1	LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyValSerLeuAspIle	20
Db	630	CTTTTGCAATTAAGTAGAGAGACTTTCAGATTTTCCTCTTGGACAGCTTTGGACATA	5711
Qy	21	AlaProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleIlyShiSproGlu	40
Db	570	GCACCAAGGTGACATCTTGACAAAGGTAGCAAGAAAGCTTTCTTAATATAAACATCCAGAG	5111
Qy	41	CysSerThrMetSerGlyGlyIlyValAlaIleGluHisLeuAlaIlySerGlnGlyAsnArgPhe	60
Db	510	TGCTCCACCAAGAGGTGGAGGCGCAATAGAACATTTGTCCAAACAGGAAATTAAGTTG	4511
Qy	61	HisPheAspIleIlyProProLeuIlyShiSalAlaIysAsnCysAspPheSerPheThrGly	80
Db	450	CATTTTGATTTCAAACTCCATCCATGCACAGCTGATATAAAATGTGATTTTCTTTTCTGGA	3911
Qy	81	LeuGlnHisValIleThrAspIlyIleIleMetIlyGlyIlySerGlnGlnGlyIleGlnIlyS	100
Db	390	CTTCAACATGTTATGATATAGACAAATATATGCAAGAAAGAAAGGAGGATATTGAGAAAG	3311
Qy	101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisPheMetAlaCys	120
Db	330	GGGCAAACTCGTCTTCACGCTGCAGACATTTGCTGCGACGTACAGACACAGTAGCCTGC	2711
Qy	121	HisIleuValIysArgThrHisArgAlaIleLeuPheCysIlySerGlnArgAspLeuLeuPro	140
Db	270	CATATGTCAAAAGAAAGCTATGCTATGCTATTCGTTTGCAAACAGAGAGCTTAATATGT	2111
Qy	141	GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg	160
Db	210	CAAGGTATGCAAGTATGCTGTTCTATCTGAGAGGTTGCAAGTAATTAATATACAAAA	1511
Qy	161	AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysAspProProArgLeu	180
Db	150	GCATTAGAGTTGTGACAAATGCACACACATGACACTGTGTGTGTCCTCCACAGACTA	91
Qy	181	CysThrAspAsnGlyIleMetIleAlaTrp	190
Db	90	TGCATCTGATTAATGGCATTATGATTCATGAG	61

RESULT 9

CK941819/c

DEFINITION	4065407	BABC 10BOV Bos taurus cDNA clone 10BOV12 F24 3' mRNA
20000	831 bp	linear
20000	831 bp	linear

sequence.

VERSION CK941819.1 GT:45456195

KEYWORDS: EST.

ORGANISM BOB taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

.

REFERENCE
AUTHORS

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 637)
Sonstegead, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay,

TITLE Production of EST from cDNA libraries derived from immunologically

JOURNAL
COMMENT

Email: tads@art1.barc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 0.000925 using options -trim at "- -trim fasta. Vector identified
 by cross match using options -mismatch 12 -minscore 18
 Plate: 12 row: F column: 24
 Seq primer: ACCGCAATCAATTTCACGAG
 High quality sequence scop: 637.

FEATURES

SOURCE

```

/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="10BOV12_F24"
/sex="Male"
/tissue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_lib="BARC 10BOV"
/note="Organ: Small Intestine; Vector: pagen-1; Site:1;
ECOR; Site 2: NotI; Equimolar amounts of mRNA extracted
from proximal jejunums of 18 and 21 wk old steers, and
distal ileums of 14 day old calves. proximal jejunum
exposed to C. oncophora for 3 and 6 weeks, and distal
ileum exposed to C. parvum for 7 days"

```

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

US-10-649-273-2_COPY_176_414 (1-239) X CK941819 (1-637)

54 A1ALYBGLNGLYAVhargrphn1vrrhavr1eluyvrtroptolewh1v1valalyvash /

Db 632 GCCAACAGGAAATAGATTGCAATTTGATTTCACGCTCCCATGCAACGTGCTAAAT 57

74 CysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlu 93

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

94 LysGIUGLUGLYLLeGIULYbGLYGLNLLeUeUSeSeTALAaLaAbpLLeALAaLaInr 11

Db 512 AAGAGGAGGTATCGAGCAGGGCAGGTCCTGTCTTCAGCTGCAGACATTCCTGTTCCG 45

114 va]G]nh:etbvmotA]aCven:et.enua]IvsaTatbvh:asvsa]at]eIeupheCva 13

[illegible][illegible]

134 LysGlnArgAspLeuProGlnAsnAlaValLeuValAlaSerGlyGlyValAla 15

Db 332 AGTACTTATATATCCGAAAGCCCTGGAAATTGTGACCAATGACACAGAGTCTTG 273
 Qy 174 LeuCySPProProxArgLeuCyThraSpanglyileMeTleAlaTrrpanglyile 193
 Db 272 CTGTGGCCGCCCCCAGACTTGGACCTGACCAAGCGCGTATGATTCATGAAATGTT 213
 Qy 194 GluArgLeuArgAlaGlyLeuGlyIleLeuHisaPpIleGlyIleArgTrrp 213
 Db 212 GAAAGACTAGCTGCTGGCTGGGCACTTTTACACACAGAAAGCATCCGCTACGACCA 153
 Qy 214 LysCySPProLeuGlyValAspIleSerIySgluValGlyGluAlaSerIleYsValPro 233
 Db 152 AAAAGTCCCTCTGGAGTATATCAAAAGAGTTGAGAGCTGTATAAAGTGCCA 93
 Qy 234 GluLeuYsMetGluIle 239
 Db 92 AGATTAAATGAAGATT 75
 RESULT 10
 LOCUS BO961028 922 bp mRNA linear EST 21-AUG-2002
 DEFINITION AGENCOURT_8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423302
 5', mRNA sequence.
 ACCESSION BO961028
 VERSION BO961028.1 GI:22376506
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 922)
 NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: csapbs@mail.nih.gov
 Tissue Procurement: DCTD/DRP/Gazdar
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LICM2603 row: d column: 15
 High quality sequence stop: 584.
 Location/Qualifiers
 1..922
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6423302"
 /issue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 18"
 /note="Organ: lung; Vector: pORF7; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCGACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIySerLeuAspIle 20
 Db 291 CTGTGGCATTAGTTCAGAGATTTCAGATTTTCCTCTTGGAAGCTTTGGACATA 350
 Qy 21 AlaProGlyAspMetLeuAspIyValAlaArgArgLeuSerLeuIleYsHisProGlu 40
 Db 351 GCACCAAGGAGATGCTTGACAAAGGTGGCAAGAGACTTTCTTAAATAAATCCAGAG 410
 Qy 41 Cys-SerThrMetSerGlyIyValAlaIleGlnHisLeuAlaYsGlnGlyIyAspArg 60
 Db 411 TGCATTCACCAAGAGGTGGGAAAGCCATGAAACCTTGGCCAAAGAAATGATT 470
 Qy 60 eHisPheAspIleYsProProLeuHisHisAlaYsAsnCyAspPheSerPheThrG 80
 Db 471 TCATTTGACATCAAACTCCCTTGATCATCTTAATAATTGATTTCTTTACTGG 530
 Qy 80 YLeuGlnHisValThrAspIyIleIleMetIySgluYsGluYsGluGlyIleGluIy 100
 Db 531 ACTTCAAACAGTTAAGTAAATAATGAAAAAGAAAAAGAGAGATTTGAGAA 590
 Qy 100 eGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAla 120
 Db 591 AGGCAAAATCCGTCTTCAGACAGACAGCACTTCGACAGTACAGACACAAATGGCATG 650
 Qy 120 HisLeuValIyAspThrHisArgAlaIleLeuPheCysGlyArgAspLeuThr 140
 Db 651 TCATCTTGAAAAGAAA-CATCGGCTATTTCTGTTTGTAAACAGAGACTTGATAC 709
 Qy 140 eGlnAsnAsnAlaValLeuValAlaSer-GlyGlyValAlaSerAspPheYrIleArgA 160
 Db 710 TCAAAATATAGCAGTACTGGTGCACTTGGGGGGTGCAGTAATCTTATATCCCG 769
 Qy 160 eGlnLeuGluIleLeuThr-AsnAlaThrGlnCysThr-LeuLeuCySPProProx 179
 Db 770 AAACCTCGGAAAAATTACAAAACCAACACGCGGCACTTTGTTGGTCCCTCCCA 829
 Qy 179 gLeuCySPThr-AspanglyIleMetIleAlaTrrp-----AsnGlyIleGluArg--L 196
 Db 830 ACTATGCACTGATTAATATGG--CATTAATGATGCTGGGGAAGGAAATTAATAATA 886
 Qy 196 eUArgAlaGlyLeuGly 201
 Db 887 CCATGCTGGCTTGGGG 903
 RESULT 11
 LOCUS AV602901/c 545 bp mRNA linear EST 27-NOV-2001
 DEFINITION AV602901 Bos taurus kidney fetus Bos taurus cDNA clone EIKI013A07
 3', mRNA sequence.
 ACCESSION AV602901
 VERSION AV602901.1 GI:9725227
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 1 (bases 1 to 545)
 Establishment of a high throughput EST sequencing system using
 poly(A) tail-removed cDNA libraries and determination of 36,000
 bovine ESTs
 Nucleic Acids Res. 29 (22), E108 (2001)
 JOURNAL MEDLINE
 PUBMED 11713328
 COMMENT Contact: Yoshikazu Sugimoto
 Animal Genetics Division
 Shikawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5541
 Fax: 81-248-25-5725

ORIGIN
 Alignment Scores:
 Pred. No.: 4.66-82 Length: 922
 Score: 799.00 Matches: 179
 Percent Similarity: 88.04% Conservative: 5
 Best Local Similarity: 85.65% Mismatches: 16
 Query Match: 54.44% Indels: 10
 DB: 5 Gaps: 3

US-10-649-273-2_copy_176_414 (1-239) x BO961028 (1-922)

Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

FEATURES

Location/Qualifiers

1..545

/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="BK1013A07"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus kidney fetus"
/note="Vector: pZL1; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"

ORIGIN

Alignment Scores:

Pred. No.: 3,18e-81 Length: 545
Score: 789.00 Matches: 147
Percent Similarity: 90.11% Conservative: 17
Best Local Similarity: 80.77% Mismatches: 17
Query Match: 63.63% Indels: 1
DB: 1 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AV602901 (1-545)

QY 57 G1YASNRGPHNHPHAPPLLELYSPROFLEUHS1S-ALALYASNCYASAPH 76
DB 544 GGAATATGATTGCAATTTGATTTCCANCTCCCAACGCGTCTGCTGAAATTTGATTT 485
QY 76 ESERPHETHG1YLEUG1NH1SVALTHRASPLYSLE1LEMETLYSLYSG1ULYSG1 96
DB 484 TTCTTTTTCGACTTCAACAGTTATTGATTAAGTATGATGCAAAAGAAANAGGA 425
QY 96 UG1Y1LG1ULYSG1ULYSG1ULYSEUSERSER1AALASPLLEALATHTVA1G1NH1 116
DB 424 AGGATGAGAGAGGGGCGAGTCTCTTCAGCTCGGACATTCGTCGCGTCCAGCA 365
QY 116 ETHMETALACYS1HSLEUVALYSA1XGTH1S1AG1A1LEUAPHECYALYSG1NAR 136
DB 364 CACCGTCCCTGCACATTCGAAAGAAACATGCTCTCTTCTGTTCTGCACAGCAG 305
QY 136 GAAPLEULEUPROG1NAPNAENAP1AVALLEUVAL1ASERG1YGLYVAL1ASERAPN 156
DB 304 AGGCTTCTTATCATGAGATGAAAGCGAGTACTGTTTATCTGAGGCGTGCAGATTA 245
QY 156 E1Y11E1ARGA1ALEUG1UL1LEU1TH1RAN1ATH1R1NG1CYSTH1R1LEUCYAPR 176
DB 244 ATATATCGAAAACCCCTCGAATTTGACCAATGCAACACAGTCCACTTGTGCCC 185
QY 176 OPRCPHARGLYCYSTH1R1ASAPENG1Y1LEMET1LEALATTPAENG1Y1LEGLUARG 196
DB 184 GCCCCCAAGACTCTGCACATGCAACGCGGTATGATTCATGATGATGATGATGATG 125
QY 196 UATG1AG1Y1LEUG1Y1LEU1H1SAP1LEGL1Y1LEA1G1Y1R1G1UPROLYCYAPR 216
DB 124 ACGGCTGCGCTTGCGCATTTTACACACAGAAAGCATTCGCTACGAAACAAATGTCC 65
QY 216 OLEUG1VALASPL1SE1R1YSG1ULYVAL1G1Y1UL1ASER1Y1LEUVAL1PROG1LEU 236
DB 64 TCTTGAGATGATATATCAAAAGAGTTGAGAGAGCTGCTATTAAGTCCCAAGATTAA 5
QY 236 Smet 237
DB 4 AATG 1

RESULT 12
LOCUS CB852881 CB852881 490 bp mRNA linear EST 22-APR-2003
DEFINITION UI-CF-FN0-af0-b-03-0-UI-s1 UI-CF-FN0 Homo sapiens cDNA clone
ACCESSION CB852881 UI-CF-FN0-af0-b-03-0-UI 3', mRNA sequence.

VERSION

CB852881.1 GI:30047942

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

1 (bases 1 to 490)
Bonald,M.P., Lennon,G. and Soares,M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

8889548

COMMENT

Contact: McCray, PB

JOURNAL

McCray Lab

MEDLINE

University of Iowa

PUBMED

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

COMMENT

Tel: 319 356 4866

JOURNAL

Fax: 319 356 7171

MEDLINE

Email: paul-mccray@uiowa.edu

PUBMED

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

COMMENT

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

JOURNAL

cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

MEDLINE

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

PUBMED

Clone Distribution: Researchers may obtain clones from Research

COMMENT

Genetics (www.research.com) or from Open Biosystems

JOURNAL

(www.openbiosystems.com).

MEDLINE

Seq primer: M13 FORWARD

PUBMED

POLYA=No.

COMMENT

Location/Qualifiers

FEATURES

1..490

SOURCE

/organism="Homo sapiens"

JOURNAL

/mol_type="mRNA"

MEDLINE

/db_xref="taxon:9606"

PUBMED

/clone="UI-CF-FN0-af0-b-03-0-UI"

COMMENT

/tissue_type="Human lung Epithelial cells"

JOURNAL

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

MEDLINE

/clone_lib="UI-CF-FN0"

PUBMED

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a

COMMENT

modified polylinker; Site 1: EcoR I; Site 2: Not I;

JOURNAL

UI-CF-FN0 is a subtracted cDNA library derived from two

MEDLINE

normalized Human lung epithelial cell libraries (EN1 and

PUBMED

DN1) The library was subtracted according to

COMMENT

Bonaldi, Lennon and Soares, Genome Research, 6:791-806,

JOURNAL

1996. For additional information, contact:

MEDLINE

bento-soares@uiowa.edu

PUBMED

tag_seq=None found"

COMMENT

ORIGIN

FEATURES

Alignment Scores:

SOURCE

Pred. No.: 1,05e-80 Length: 490

JOURNAL

Score: 784.00 Matches: 148

MEDLINE

Percent Similarity: 100.00% Conservative: 0

PUBMED

Best Local Similarity: 100.00% Mismatches: 0

COMMENT

Query Match: 63.23% Indels: 0

JOURNAL

DB: 6 Gaps: 0

MEDLINE

US-10-649-273-2_COPY_176_414 (1-239) x CB852881 (1-490)

QY 69 H1SH1SALALYASNCYASAPPHSESPHETH1R1NG1Y1LEUG1NH1SVALTH1R1APPLYS1LE 88
DB 40 CATCATGCTGAAAATTTGATTTTCTTTTACTGAGCTTCAACACGTTACTGATTAATA 99
QY 89 11EMETLYSLYSG1ULYSG1ULYSG1ULY1LEGLU1Y1LEU1Y1LEU1SER1A1A1A 108
DB 100 ATATATGAAAAGGAAAAGGAAAGGATTTGAGAAAGGCAAAATCTGCTTCAAGCA 159
QY 109 AAP11EAL1ATH1R1VAL1G1NH1STH1R1MET1ALACYS1HSLEUVAL1YSA1GTH1R1SA1X 128
DB 160 GACATTGCTGCCACAGTACGACACAAATGCGATGTCATCTTGTGAAGAAACACATCGG 219
QY 129 A1A11E1LEUPH1CYALYSG1NAR1AP1LEU1PROG1NAPNAPN1AVALLEUVAL1A1A 148

Db 220 GCATTCGTTGTTGTAAGCAGAGACTTGTTACTCAAAATATATGCACTACTGTTGGC 279

QY 149 SerGIYGIYValAlaSerAspPheTYrIleArgAlaLeuGluIleuThrAspAla 168

Db 280 TCTGGTGTGTGCGAAGTAACTTCTATATCCCAAGCTCTGGAAATTTTAAACAAGCA 339

QY 169 ThrGlnCYthrLeuLeuCYsPheProPheArgLeuCYthrAspAsnGlyIleMetIle 188

Db 340 ACACAGTCACTTGTGTGTCTCTCCACAGACTATGACATGATATGCACTTATGATT 399

QY 189 AlaTTPanGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGly 208

Db 400 GCATGGAATGGATATGAAAGACTACGTGCTGGCATTTTACATGACATAGAGGC 459

QY 209 IleArgTYrGluProLeuCYsPhePro 216

Db 460 ATCCGCTATGAAACCAAAATGTCCC 483

RESULT 13

CF257246 822 bp mRNA linear EST 07-AUG-2003

LOCUS phao08_g02 PHA-activated splenocytes Gallus gallus cdna, mRNA

DEFINITION sequence.

ACCESSION CF257246 GI:33490501

VERSION CF257246.1

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 822)

Wiltzell, H., Bed'Hom, B., Morin, V., Young, J.R., Whitaker, C.J., Chausse, A.M. and Zoorob, R.

A collection of chicken ESTs from activated immune cells

TITLE Unpublished (2003)

JOURNAL Contact: Zoorob R

COMMENT CNRS 1983

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France

Tel: 33 1 49 58 35 00

Fax: 33 1 49 58 33 81

Email: zoorob@vjf.cnrs.fr.

FEATURES

source location/Qualifiers

1..822

/organism="Gallus gallus"

/mol_type="mRNA"

/db_xref="taxon:9031"

/cell_type="Splenocytes"

/clone_lib="PHA-activated splenocytes"

/note="Vector: pTIR1EX2"

ORIGIN

Alignment Scores:

Pred. No.: 2.85e-80 Length: 822

Score: 783.00 Matches: 150

Percent Similarity: 79.57% Conservative: 37

Best Local Similarity: 63.83% Mismatches: 48

Query Match: 63.15% Indels: 1

DB: 7 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x CF257246 (1-822)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db 119 ATCTTGGAGTAGACAGCAGAGTTTCAGATTCTTCTGCTTGACAGTCCATAGATCA 178

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

Db 179 GCACCAAGTATGCAATGTTCATACCTACCAAGAGGCTCTTTAGTGAAGCACCAGGAG 238

QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

Db 239 TGCACAGCAGATGCGCGGGGGAAGCATTAGACACCTGTGCTCAACCGAGACTGGCAA 298

QY 61 HisPheAspIleLysProPheLeuHisAlaLysAsnCYsAspPheSerPheThrGly 80

Db 299 CACTACACTTCAAGACTCTCCATGCAACAGTATGCTATGATTTTCTTCTCCGGA 358

QY 81 LeuGlnHisValThrAspLysIleIleMetLysGluLysGluGluGlyIleGluLys 100

Db 359 CTTCAAGGCTGTCAACAAGCATTCTTCAGAAAGAAAAGAAAGAGATTCAGAA 418

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120

Db 419 GGGGAATCTCTGTCTCGCTTAAGACATGCTGCTGTCCACAGCAGTATGCTGCT 478

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCYsLysGlnArgAspLeuLeuPro 140

Db 479 CATATTATCCAGCGGACACACGAGCATCTTCTTGCAATGAAAACAGCATATTATTA 538

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTYrIleArgArg 160

Db 539 CCAAAACATGCAACTGTGTTGATACAGAGAGCTTGCAAGTATATCATCAGAAA 598

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCYthrLeuLeuCYsPheProPheArgLeu 180

Db 599 GGACTCAGACTCTGCGCAATGCAACGGTTTGTCTTGTCTCTCTCCCAAGGCTG 658

QY 181 CYthrAspAsnGlyIleMetIleAlaTPanGlyIleGluArgLeuArgAlaGlyLeu 200

Db 659 TGCACCATTAATGTGTATGATGTGATGCAATGCAATGCAATGCAATGCAATGCAATG 718

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTYrGluProLysCYsPheLeuGlyValAsp 220

Db 719 GGAATTTATACAGTACTATGATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 778

QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGluLeu 235

Db 779 ATTTCCAAAGAGTTGAA-GAGGATTCATCAAAAGTCCCAAGACTA 822

RESULT 14

BUI27463 866 bp mRNA linear EST 25-NOV-2002

LOCUS 603114407F1 CSBQCHL20 Gallus gallus cdna clone CHEST6m2 5', mRNA

DEFINITION sequence.

ACCESSION BUI27463 GI:25338728

VERSION BUI27463.1

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 866)

Bordman, P.B., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fond, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

A Comprehensive Collection of Chicken cDNAs

TITLE Curr. Biol. 12 (22), 1965-1969 (2002)

JOURNAL 22335534

MEDLINE 12445392

PUBMED Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source location/Qualifiers

1..866

/organism="Gallus gallus"

/mol_type="mRNA"

/strain="White Leghorn, Hisex"

/db_xref="taxon:9031"

/clone="CHST6m2"

```

/dev_stage="36"
/lab_host="DHI09"
/clone_lib="CSECHL20"
/note="Oxyn: limbs; Vector: pbluescript II KS(+); Site 1
EcoRI; Site 2: NotI. Modification of pbluescript II KS(+)
Stratagene) vector to accommodate cDNA produced with the
l-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Pl., et al. U.S. Patent #
6,387,624). Cut pbluescript II KS(+) with NotI and EcoRI.
ligate in double stranded adaptor containing BspGI and
BamHI sites [5'ggcgcgcygcacgcccgcatctgaataaag]
[5'aattcttttctgcgatccggcctgcgaoc]"

```

ORIGIN	
Alignment Scores:	
Pred. No.:	9,58e-77
Score:	753.00
Percent Similarity:	78.81%
Best Local Similarity:	63.14%
Query Match:	60.73%
DB:	5
Length:	866
Matches:	149
Conservative:	37
Mismatches:	50
Indels:	2
Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) X BU127463 (1-866)

Qy	1	LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGluYyssSerLeuAspIle	20
		::: :::	
Db	25	ATCTGGCAGTGTACGACGAGGAGTTCCATGATTTCTCTGCTTGACGACGTCATGATATA	84
Qy	21	AlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleIysHsIspProGlu	40
		::: :::	
Db	85	GCACCGATGACATCTGGATTAAGGTGACGAAGGCTCTCTTTAGTGAAGCACCCGGAG	144

41 CysserThmetSecdlygylsaaalleguhsisneualaysglmglyanarphe 60
 145 TGCCACGGCATGGCTGGGGGAGGCGCATAGACACCTGGCTCGAACCGGAGCTGGCNA 204

Qy 61 HisSheApriLelysProPLeuHiShiSaLaLyasncYsaSPheSerPhetHclY 80
:
:
b 205 CAGTACACTTCGACCTCCATGCACAAGATCGTAAGTAATTTCTTTCTCCGA 2644
:
:
:

Oy 81 LeuGlnHisValThrAspIysAllelleMetLysLysGluLysGluGluGlyIleGluLys 100
||||| :||| | ||||| ||||| ||||| ||||| :||| :|||
Db 265 CTTCAAGACCTGTGTACAACACCATTCTTCAGAAAGAAAAAAGAAGATATTCAAAG 324

Qy	101	GLGlnileuSeSerAlaAlaSpil	leAlaAlaThrValGlnHs	sthrMetAlaCys	120
Db	325	GGGAAATCTGTCTTCGTTAAGACATCGCTGCTGCACAGCATGTAGTGGCTCT			384

Qy 121 HisleuValIysargThrHisargAlaIleuPheCysIysGlnargAspIeuLeuPro 140

Cy 141 GlnAsnAsnaIaValLeuValAlaSerGlyVcIyVaLaAlaSerAsnphetyrIleargarg 160
445 CCAAAACCGCAGCCATCTTATTGGAGCCATTCCTCATCTAATCTATCTCTGCAGA :	

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgGlu 180

161 Cythrpaapnglytlemetleatrpanglytlegluarleuargalaglyleu 200

201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220

221 Ilse Se er ly ec lu va lg yl gi ua se ri le ys va l pr og in le u ys 236

DB 683 AATTCCAAAAGAGTTGAAGAGATTCATCAGAGTCCAAAGCTAAGS 730 .

RESULT 15	CK365185	LOCUS	DEFINITION
	CK365185	736 bp	mRNA
			linear
			EST 23-DEC-2003
			AGENCOURT_17157291 NIH MGC 221 Rattus norvegicus cDNA clone
			IMAGE:7097620 5', mRNA sequence.

```

VERSION      CK365185.1  GI:40331120
KEYWORDS
SOURCE       Rattus norvegicus (Norway rat)
EST.

```

ORGANISM
Rattus norvegicus
Bakaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE	1 (bases 1 to 736)
AUTHORS	NIH-MGC http://mgc.ncbi.nlm.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC).
JOURNAL	Unpublished (1999)
COMMENT	Contact: Daniela S. Gerhard, Ph.D. gerhard@nigms.nih.gov

National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs

CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLAM14944 row: d column:
High quality sequence stop: 736.

FEATURES	Location/Qualifiers
source	1. .736

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7097620"
/tissue_type="lung, pooled"
/lab_host="DH10B Tona"
/clone_1b="NIH MGC 231"
/note="Organ: lung; Vector: pExprres-1; Site 1: EcoRV;
Site 2: NotI; RNA obtained from pooled lung tissue from a
mix of male and female animals at 8 wk old. Tissues were
snap-frozen and kept at -80C for two days before RNA
extraction and purification (TRI-reagent method). cDNA was
primed using oligo-dT primer:
5'-pGACTGATTCAGATCCGACGCGCCGCC(C)25-3' and cloned into
the EcoRV/NotI sites of pExprres-1. Size-selection >1.4kb
resulted in an average insert size of 2.3 kb. This primary
library is not normalized (normalized primary library is
NIH MGC 232) and was constructed by Express Genomics
(Fredrick, MD). Note: this is a NIH MGC library."

```

ORIGIN

Alignment Scores:	
Pred. No.:	2,22e-76
Score:	749.00
Percent Similarity:	91.02%
Best Local Similarity:	84.43%
Query Match:	60.40%
DB:	7
Length:	736
Matches:	141
Conservative:	111
Mismatches:	15
Indels:	0
Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) X CK365185 (1-736)

Oy 73 AsnGAspPheSerPheThrGluLeuGlnHisValThrAspLysIleIleuMetLysLys 92
 Db 1 AACTGTAATTTTCTTTACGGGACTTCAACATGTCACCGGATTAAGGTAATATACACACAG 60

93 GlnGysGlnGlnGlyIleGlnGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 112
QY
61 GAAAAAGAGAAAGGCATTGAGAAAGGGCAATCCTGCATCAGCCGAGACATTGCTCT 120
DB

113 ThrValGlnHisThrMetAlaCysHisIleuValTyrArgThrHisArgAlaIleLeuPhe 132

```
Db 121 GCGGTACAGACGCGACAGCGCTGCCACTTGCGAAAAGACACATCGTGTCTGTT 180
QY 133 CybelysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 152
Db 181 TGGCAGCAGAAAAATTGCTATCTCCAGCTAACGCGATATTAGTGTGTGAGGTGTT 240
QY 153 AlaserAspPheTyrLeuArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 172
Db 241 GCAGTAACTTGTATCATCCGAGAGCATTTGGAATTGTAGCAAAATGCACACATGCACT 300
QY 173 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGly 192
Db 301 TTGTTGTGTCCCTCCGAGACTGTGCACTGACATGTGATCATGATTCATGCAATGGA 360
QY 193 IleGluArgLeuAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 212
Db 361 ATTGAAGATTACGTCTGCTGTGGCATTTTACATGATGTAGAGACATCGATACGAA 420
QY 213 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 232
Db 421 CCAAAATGTCTCTCGGATATGACATATCCAGAGAGTTGCGAGAGCTGCCATAAAGTA 480
QY 233 ProGlnLeuLysMetGluIle 239
Db 481 CCAAGATTAAAAATGACACTT 501
```

Search completed: February 16, 2005, 21:04:55
Job time : 2847.68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 10:46:17 ; Search time 40.5261 Seconds
(without alignments)
3019.954 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVQGVSDPFLIGKSLDI.....DISKVGESAIKVPOLKMEI 239

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	414	2	096EV9 homo sapien
2	1213.5	97.9	439	2	09H4B0
3	1086	87.6	414	2	06PEB4
4	1079	87.0	414	2	08BLB6
5	1071	86.4	414	2	09D0N0
6	983	79.3	467	2	06AYN7
7	960	77.4	364	2	096NH5
8	751	60.6	404	2	08JFW3
9	747	60.2	404	2	08JFW3
10	361.5	29.2	401	2	07Q9I8
11	332.5	26.8	409	2	0960S6
12	332.5	26.8	409	2	09VMD6
13	324	26.1	480	2	022145
14	299.5	24.2	255	2	0677H2
15	262.5	21.2	360	2	092LH8
16	257	20.7	346	2	07VXN4
17	254	20.5	335	2	07VXN4
18	253	20.4	362	2	098X16
19	252.5	20.4	340	2	06FCR9
20	252	20.3	346	2	07W6E8
21	251	20.2	346	2	06ND54
22	246.5	19.9	389	2	06NWX4
23	245	19.8	387	2	068XR3
24	244	19.7	366	2	08UC47
25	244	19.7	366	2	07CWJ8
26	241.5	19.5	359	2	08YJBI
27	241.5	19.5	359	2	08FYI5
28	239	19.4	241	2	07CJL7
29	239	19.3	251	2	093FJ2
30	238	19.2	387	1	09E8R8
31	237.5	19.2	367	2	09ABZ9

32	236.5	19.1	357	2	089WV1	089WV1 bradyrhizob
33	234.5	18.9	344	2	07PAG7	07PAG7 rickettsia
34	233.5	18.8	339	2	06LVI0	06LVI0 photobacter
35	231	18.6	341	2	07NIB3	07NIB3 chromobacte
36	231	18.6	341	2	091SV7	091SV7 pseudomonas
37	231	18.6	364	2	06GIR3	06GIR3 bartonella
38	230.5	18.6	353	2	092UK6	092UK6 rickettsia
39	228.5	18.4	344	2	07VQO9	07VQO9 candidatus
40	224	18.1	337	2	082XN2	082XN2 nitrosomonas
41	224	18.1	342	1	GCP HAEIN	P43764 haemophilus
42	223.5	18.0	344	2	065EP0	065EP0 mannheimia
43	221.5	17.9	347	2	0291S3	0291S3 archaeoglob
44	219.5	17.7	347	2	08XX97	08XX97 ralsotonia s
45	217	17.5	341	2	08ES16	08ES16 oceanobacill

ALIGNMENTS

RESULT 1
ID 096EV9 PRELIMINARY; PRT; 414 AA.
AC 096EV9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE OSCEPPL protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteleia; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22188257; PubMed=12477932; DOI=10.1073/pnas.242638999;
RA Strausberg R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Maltby S.J.,
RA Boeak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX Strausberg R.;
RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
EMBL: BC011904; AAH11904.1; --
DR MEROPS; M22.004; --
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00905; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Ostalg.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF004537; Ostalgic_Pctdc; 1.
DR PRINTS; PR00789; OSTALOPTASR.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp_1.
SQ SEQUENCE 414 AA; 45122 MW; AS36B333F5C6B8DD CRC64;

Query Match 100.0%; Score 1240; DB 2; Length 414;
 Best Local Similarity 100.0%; Pred. No. 5,6e-101;
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQVSDPFLIGKSLDIAPGMDLKVARRLSLTKHPECSTMSGKAIIEHLAKQGNRF 60
 DB 176 LIALVQVSDPFLIGKSLDIAPGMDLKVARRLSLTKHPECSTMSGKAIIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAKNCDPFTGLQHVTDKIMKKEKEGEGILSSADIAATVQHTMAC 120
 DB 236 HFDIKPPLHAKNCDPFTGLQHVTDKIMKKEKEGEGILSSADIAATVQHTMAC 295
 QY 121 HLKRTTRAILFCQKORDLLPNNNAVLVAGSVAFYIRRALEILTNAQCTLLCPPL 180
 DB 296 HLKRTTRAILFCQKORDLLPNNNAVLVAGSVAFYIRRALEILTNAQCTLLCPPL 355
 QY 181 CTDNGIMIAMNGIERLRAGLILHDIGIRYEPKCPGLGVDSKEVGASIKVPOLKMEI 239
 DB 356 CTDNGIMIAMNGIERLRAGLILHDIGIRYEPKCPGLGVDSKEVGASIKVPOLKMEI 414

RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.
 AC 09H4B0
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Putative sialoglycoprotease type 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Chen J.M., Fortunato M., Barrett A.J.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ295148; CAC14666.1; -.
 DR MEROPS; M22.004; -.
 DR Genew; HGNC:23075; OSGEPL1.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PRINTS; PR00789; OSIALGLC_ptide; 1.
 DR PRODOM; PD002367; Peptidase_M22; 1.
 DR TrEMBL; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 439 AA; 48040 MW; 44849372C784E41F CRC64;

Query Match 97.9%; Score 1213.5; DB 2; Length 439;
 Best Local Similarity 90.2%; Pred. No. 1.3e-98;
 Matches 239; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 LIALVQVSDPFLIGKSLDIAPGMDLKVARRLSLTKHPECSTMSGKAIIEHLAKQGNRF 60
 DB 176 LIALVQVSDPFLIGKSLDIAPGMDLKVARRLSLTKHPECSTMSGKAIIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAKNCDPFTGLQHVTDKIMKKEKEGEGILSSADIAATVQHTMAC 120
 DB 236 HFDIKPPLHAKNCDPFTGLQHVTDKIMKKEKEGEGILSSADIAATVQHTMAC 295
 QY 99 ---EKGOILSSADIAATVQHTMACHLVKTTRAILFCQKORDLLPNNNAVLVAGSVASN 155
 DB 296 CRVEKGOILSSADIAATVQHTMACHLVKTTRAILFCQKORDLLPNNNAVLVAGSVASN 355
 QY 156 FYIRRALEILTNAQCTLLCPPLCTDNGIMIAMNGIERLRAGLILHDIGIRYEPKC 215

DB 356 FYIRRALEILTNAQCTLLCPPLCTDNGIMIAMNGIERLRAGLILHDIGIRYEPKC 415
 QY 216 PLGVDSKEVGASIKVPOLKMEI 239
 DB 416 PLGVDSKEVGASIKVPOLKMEI 439

RESULT 3

Q6PEB4 PRELIMINARY; PRT; 414 AA.
 AC 06PEB4;
 DT 05-JUL-2004 (TREMblrel. 27, Created)
 DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=1247932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Dege J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stachenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Lounellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska J., Smalins D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RA Strausberg R.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC058172; AAH58172.1; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PRINTS; PR00789; OSIALGLC_ptide; 1.
 DR PRODOM; PD002367; Peptidase_M22; 1.
 DR TrEMBL; TIGR00329; gcp; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBCAE CRC64;

Query Match 87.6%; Score 1086; DB 2; Length 414;
 Best Local Similarity 85.8%; Pred. No. 2.2e-87;
 Matches 205; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 LIALVQVSDPFLIGKSLDIAPGMDLKVARRLSLTKHPECSTMSGKAIIEHLAKQGNRF 60
 DB 176 LIALVQVSDPFLIGKSLDIAPGMDLKVARRLSLTKHPECSTMSGKAIIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAKNCDPFTGLQHVTDKIMKKEKEGEGILSSADIAATVQHTMAC 120

Db 236 HFTINPMPONAKNCDSPFTGLQHTDKLITNKEKEGIEKGQILSSADIAAHOHTAC 295
 Qy 121 HLVRKTRHAILFCOKRDLIPONNAVVASGVASNFYIRALEITLNATQCTLLCPPL 180
 Db 296 HLAKRTHRAILFCOKNLSPPANNAVVGVSASNLIRKALEIVANAQCTLLCPPL 355
 Qy 181 CTDNGIMIMANGIERLRAGLIGLIDIEGIRYEPKCPGLGVDSISKEYGEASIKVPOLKMEI 239
 Db 356 CTDNGIMIMANGIERLRAGLIGLIDIEGIRYEPKCPGLGVDSISKEYGEASIKVPOLKMEI 414
 RESULT 4
 Q8BLB6 PRELIMINARY; PRT; 414 AA.
 AC Q8BLB6;
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
 DE enriched library, clone:B230219017 product:similar to PUTATIVE
 DE SIALOGLYCOPROTEASE TYPE 2.
 GN Name-Osegep11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RA The RIKEN Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Saeki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishi Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujisaki S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanabe K.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 [6]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RA Adachi Y., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirokane T.,
 RA Hori F., Imotani K., Ishi Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Koyama S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akaiwa S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK045669; BAC32450.1; -
 DR MEROBS; M22.004; -
 DR MGD; MGI:1919335; Osegep11.
 DR GO; GO:0008450; F-0-sialoglycoprotein endopeptidase activity, IEA.
 DR GO; GO:0008233; F-peptidase activity, IEA.
 DR GO; GO:0008270; F-zinc ion binding, IEA.
 DR GO; GO:0006508; P-proteolysis and peptidolysis, IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PRINTS; PIRSF004537; Osialglc.pptds; 1.
 DR PRODOM; PD002367; Peptidase_M22; 1.
 DR TIGRPFAMs; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;
 Query Match 87.0%; Score 1079; DB 2; Length 414;
 Best Local Similarity 85.8%; Pred. No. 9,1e-87;
 Matches 205; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
 Qy 1 LIALVGVSDPFLIGKSLDIPAGMDKVARRLSLIKHEPCSTMSGKALIEHLAKQGNRP 60
 Db 176 LIALVGVSDPFLIGKSLDIPAGMDKVARRLSLIKHEPCSTMSGKALIEHLAKQGNRP 235
 Qy 61 HEDIKRPLHAAKCDSPFTGLQHTDKLITNKEKEGIEKGQILSSADIAAHOHTAC 120
 Db 236 HFTINPMPONAKNCDSPFTGLQHTDKLITNKEKEGIEKGQILSSADIAAHOHTAC 295
 Qy 121 HLVRKTRHAILFCOKRDLIPONNAVVASGVASNFYIRALEITLNATQCTLLCPPL 180
 Db 296 HLAKRTHRAILFCOKNLSPPANNAVVGVSASNLIRKALEIVANAQCTLLCPPL 355
 Qy 181 CTDNGIMIMANGIERLRAGLIGLIDIEGIRYEPKCPGLGVDSISKEYGEASIKVPOLKMEI 239
 Db 356 CTDNGIMIMANGIERLRAGLIGLIDIEGIRYEPKCPGLGVDSISKEYGEASIKVPOLKMEI 414
 RESULT 5
 Q9DON0 PRELIMINARY; PRT; 414 AA.
 ID Q9DON0;
 AC Q9DON0;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 DE enriched library, clone:2610001M19 product:similar to PUTATIVE
 DE SIALOGLYCOPROTEASE TYPE 2.
 GN Name-Osegep11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo body;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RA The FANTOM Consortium,
the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
Kanno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subcloning of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20530911; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
Kanno H., Akiyama J., Nishi K., Kitsuura T., Tashiro H., Itoh M.,
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
Okazaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsunura S., Kawai J.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multichannel sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
Imotoai K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ono M.,
Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
Tajima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK011265; BAB37506.1; -
DR MEROPS; M22.004; -
DR MGD; MGI:191335; Osegep1.
DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00905; Peptidase_M22.
DR InterPro; IPR00180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRPMs; TIGR00329; gcp; 1.
KW Protease.
SQ SSQENSE 414 AA; 44999 MM; 999BC689944DB24 CRC64;
Query Match 86.4%; Score 1071; DB 2; Length 414;
Best Local Similarity 85.4%; Pred. No. 4,6e-86;
Matches 204; Conservative 14; Mismatches 21; Indels 0; Gaps 0;

QY 1 LIALVGVSDPFLILGSLDIAPDMDKYARRSLIKHPECSTMSGKATIEHLAKGNRF 60
Db 176 LIALVGVSDPFLILGSLDIAPDMDKYARRSLIKHPECSTMSGKATIEHLAKGNRF 235
QY 61 HPDIKPEPLHARNCDFSPFGLQHTVDKIIMKEKEBGEIKGQILSSAADIAATVQHTMAC 120
Db 236 HFTINPQMANKNCDFSPFGLQHTVDKIIMKEKEBGEIKGQILSSAADIAATVQHTMAC 295
QY 121 HLKRTTRAILPCKQDILLPNNNAVIVAGGVASNFYIRALAILTNATQCTILCPPL 180
Db 296 HLKRTTRAILPCKQDILLPNNNAVIVAGGVASNFYIRALAILTNATQCTILCPPL 355
QY 181 CTNDGIMIMWNGIERPAGLIGLHDEGIRYEPKCPGLVDISKVGBASIKVPOLKMEI 239
Db 356 CTNDGIMIMWNGIERPAGLIGLHDEGIRYEPKCPGLVDISKVGBASIKVPOLKMEI 414
RESULT 6
ID 06A9N7 PRELIMINARY; PRT; 467 AA.
AC 06A9N7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,
Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gamarale P.H.,
Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
Jones S.J., Maiz M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Director MGC Project;
RT Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC078974; AAH78974.1; -
DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00905; Peptidase_M22.
DR InterPro; IPR00180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRPMs; TIGR00329; gcp; 1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
KW Hypothetical protein.

FT NON_TER 467 467
 SQ SEQUENCE 467 AA; 50799 MW; 474E18B1959B8AC0 CRC64;
 Query Match 79.3%; Score 983; DB 2; Length 467;
 Best Local Similarity 87.9%; Pred. No. 3.1e-78;
 Matches 186; Conservative 9; Mismatches 17; Indels 0; Gaps 0;
 QY 1 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 60
 DB 176 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 235
 QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMAC 120
 DB 236 HFTINPMQAKNCDPSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAAATVQHTMAC 295
 QY 121 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALAILTNATQCTLLCPPRL 180
 DB 296 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALAILTNATQCTLLCPPRL 355
 QY 181 CTDNGIMIAMNGIERLRAGLILHDIEGIRYEPK 214
 DB 356 CTDNGIMIAMNGIERLRAGLILHDIEGIRYEPK 389
 RESULT 7
 Q96NH5 PRELIMINARY; PRT; 364 AA.
 AC Q96NH5; 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ30879.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RC PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa B., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Ninomiya K., Itohashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Nishibashi K., Yuki H., Oshima A., Sasaki N., Aotsuma S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki M., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukunishi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Ohtsuki R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki K.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 cDNAs.";
 PT Nat. Genet. 36:40-45(2004).
 DR EMBL AK055441, BAB70923.1; -
 DR MEROPS; M22.004; -.

DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialyl.
 DR Pfam; PF00814; Peptidase_M22_1.
 DR PIRSF; PIRSF004537; Osialylc_ptds; 1.
 DR PRINTS; PR00789; OSIALOPRASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 364 AA; 39528 MW; E0E605A07D0EC3D6 CRC64;
 Query Match 77.4%; Score 960; DB 2; Length 364;
 Best Local Similarity 98.4%; Pred. No. 2.5e-76;
 Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 60
 DB 176 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 235
 QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMAC 120
 DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMAC 295
 QY 121 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALAILTNATQCTLLCPPRL 180
 DB 296 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALAILTNATQCTLLCPPRL 355
 QY 181 CTDNGIMIAM 189
 DB 356 CTDNGIMIAM 364
 RESULT 8
 Q8JFW3 PRELIMINARY; PRT; 404 AA.
 AC Q8JFW3; 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE SI-42211013.4 (Novel glycoprotease).
 GN Name=d72B14.6;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Babage A.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL591593; CAD43471.1; -
 DR MEROPS; M22.004; -
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009180; Pept_M22_Osialyl.
 DR InterPro; IPR009180; Pept_M22_Osialyl.
 DR Pfam; PF00814; Peptidase_M22_1.
 DR PIRSF; PIRSF004537; Osialylc_ptds; 1.
 DR PRINTS; PR00789; OSIALOPRASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737E35 CRC64;
 Query Match 60.6%; Score 751; DB 2; Length 404;
 Best Local Similarity 61.6%; Pred. No. 7.9e-58;
 Matches 146; Conservative 32; Mismatches 59; Indels 0; Gaps 0;

QY 1 L1ALVGVSDPFLGKSLDIPAGDMLDKVARRLSLIKPECSMTSGKAIEHLAKGNRF 60
 DB 164 L1ALAKGIDFLLGQTLDBAAGDTLDKIARLRLSRNPECGTLLSGGAIERLAKEGDRL 223
 QY 61 HFDIKPPLHAKNCDPSTGLOHTVDKIMKKKEEGEIKGQILSSADIATVOHTMAC 120
 DB 224 AFHFIISPWGQYDNCFSFAGLRTOITGAIKKKEKEGVEAGQPLSCVKDIAAASQHTVAS 283
 QY 121 HLVRTRHAILFCQKQDILLPQNNANVLVAGSVANFYIRRALEITLMTAQTCTLCPPRL 180
 DB 284 HLARTRHAILFCCKSKGLLPQNPFLIVSGVANSBYIRQILKITDADGHLCPSPRF 343
 QY 181 CTDNGIMIANNGIERLAKGILHIDIEGIRYEPKCPPLGVDISKEVGEASIKVPOLKM 237
 DB 344 CTDNGVMIANNGIERLKQKGLISSESVSEYEPKAPLGLDITSVEAKAIKVPKLDL 400

RESULT 9

Q8JFR7 PRELIMINARY; PRT; 404 AA.
 AC 08JFR7
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE SI:d72B14.6 (Novel glycoprotease).
 GN Name=SI:d72B14.6;
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark G.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL672217; CAD3443.1; -.
 DR MEROPS; M22.004; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008223; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase M22.
 DR InterPro; IPR009180; Pept M22 Ostalgl.
 DR Pfam; PF00814; Peptidase_M22_1.
 DR PIRSF; PIRSF004537; Ostalgl.pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KM PROTEASE
 SQ SEQUENCE 404 AA; 44027 MW; 6F59653A651860F CRC64;

Query Match 60.2%; Score 747; DB 2; Length 404;
 Best local similarity 61.2%; Pred. No. 1.8e-57;
 Matches 145; Conservative 33; Mismatches 59; Indels 0; Gaps 0;

QY 1 L1ALVGVSDPFLGKSLDIPAGDMLDKVARRLSLIKPECSMTSGKAIEHLAKGNRF 60
 DB 164 L1ALAKGIDFLLGQTLDBAAGDTLDKIARLRLSRNPECGTLLSGGAIERLAKEGDRL 223
 QY 61 HFDIKPPLHAKNCDPSTGLOHTVDKIMKKKEEGEIKGQILSSADIATVOHTMAC 120
 DB 224 AFHFIISPWGQYDNCFSFAGLRTOITGAIKKKEKEGVEAGQPLSCVKDIAAASQHTVAS 283
 QY 121 HLVRTRHAILFCQKQDILLPQNNANVLVAGSVANFYIRRALEITLMTAQTCTLCPPRL 180
 DB 284 HLARTRHAILFCCKSKGLLPQNPFLIVSGVANSBYIRQILKITDADGHLCPSPRF 343
 QY 181 CTDNGIMIANNGIERLAKGILHIDIEGIRYEPKCPPLGVDISKEVGEASIKVPOLKM 237
 DB 344 CTDNGVMIANNGIERLKQKGLISSESVSEYEPKAPLGLDITSVEAKAIKVPKLDL 400

RESULT 10

Q7Q918 PRELIMINARY; PRT; 401 AA.
 ID 07Q918
 AC 07Q918
 DT 01-MAR-2004 (Tremblrel. 26, Created)
 DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE AGCP14990 (Fragment).
 GN Name=agcG46164; ORFNames=ENSANG0000007922;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
 OX NCBI_Taxid=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.

CC EMBL; AAB01008900; EAA09387.1; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase M22.
 DR InterPro; IPR009180; Pept M22 Ostalgl.
 DR Pfam; PF00814; Peptidase_M22_1.
 DR PIRSF; PIRSF004537; Ostalgl.pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 FT NON TER 1
 SQ SEQUENCE 401 AA; 43849 MW; CC9426723D3FDAF1 CRC64;

Query Match 29.2%; Score 361.5; DB 2; Length 401;
 Best local similarity 36.3%; Pred. No. 1.9e-23;
 Matches 90; Conservative 44; Mismatches 103; Indels 11; Gaps 4;

QY 1 L1ALVGVSDPFLGKSLDIPAGDMLDKVARRLSLIKPECSMTSGKAIEHLAKGNRF 57
 DB 147 L1LVFVSTARFLGTLDDAPEALDKIARLRLSRNPECGTLLSGGAIERLAKGNRF 206
 QY 58 --NRHFDIKPPLHAKNCDPSTGLOHTVDKIMKKKEEGEIKGQILSSADIATVOHTMAC 115
 DB 207 DTSAYEPL--PLSKTRDQCFSPAGLRTOITGAIKKKEKEGVEAGQPLSCVKDIAAASQHTVAS 264
 QY 116 HTMAACHLVKTRHAILFCQKQDILLPQNNANVLVAGSVANFYIRRALEITLMTAQTCTLCPPRL 173
 DB 265 KQVTRHAILFCCKSKGLLPQNPFLIVSGVANSBYIRQILKITDADGHLCPSPRF 324
 QY 174 LCPPLCTDNGIMIANNGIERLAKGILHIDIEGIRYEPKCPPLGVDISKEVGEASIKVPOLKM 231
 DB 325 YPPPKLCTDNGIMIANNGIERLAKGILHIDIEGIRYEPKCPPLGVDISKEVGEASIKVPOLKM 384
 QY 232 VPQLKMEI 239
 DB 385 CKWAKVDI 392

RESULT 11

Q96086 PRELIMINARY; PRT; 323 AA.
 ID 096086
 AC 096086
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE LD37221P.
 GN ORFNames=CG14231;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;

[1]
RA SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Apbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Paclele J., Paragas V., Park S., Pounanavong S., Wan K.,
RA Yu C., Lewis S.B., Rubin G.M., Celniker S.,
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051882; AAK93306.1; -
DR FlyBase: FBgn0031060; CG14231.
DR GO: GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase M22.
DR InterPro: IPR009180; Pept M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_ptds; 1.
DR PRINTS: PR00789; OSIALOPTASE.
SQ SEQUENCE 323 AA; 35828 MW; BEDB54D9A2BF35A CRC64;

Query Match 26.8%; Score 332.5; DB 2; Length 323;
Best Local Similarity 35.9%; Pred. No. 5.3e-21;
Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

QY 2 LALVQGVSDPLILGKSLDIAFGDMLDKVARSLILKPECSGSKAIEHLAK-QGNRF 60
DB 81 LVVANGGRGLTLGQTLDDARGEAFDKIGRLRLHLPEYLMNGRAIEHNAQLASPL 140
QY 61 HEDIKPEPLHAKNCDSPFTQVHTDKIMKEKEGIEKQILLSSADIAATVQHTMAC 120
DB 141 AVEFPLPLAQQNCFNSFAGIKNNNSFRAIRARERAEPTPGVIVNYGDFCGILRSVR 200
QY 121 HLYKRTIRALFC--KQDRLPQNNATLVAGGVASNYTIRALEILTNAQCTILCP 178
DB 201 HLMKRTQRALEYCYLPHROGFDPPTLVMSGVANDAIYANIETHLAQYGRSFRPSK 260
QY 179 RLCTDNGIMIANNGIERLRAGLILHIDE-GIRYEPKPCPLVDISKVEGA 228
DB 261 RYSDNGVMIAHMGVEQL-----LQDKKASTRTDYD--SIDIQSGAGFA 302

RESULT 12
QYVMD6 PRELIMINARY; PRT; 409 AA.
AC QYVMD6; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG14231-PA.
GN ORFNames=CG14231;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; Pubmed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.B., Holt R.A., Bays C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.B., Li P.W., Hoskins R.A., Galie R.P.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter R.G., Helt G., Nelson C.R., Gabot G.L.,
RA Abril J.F., Apbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Beasly A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokhov D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis J.C., Bustam D.A., Butler H., Cauden B., Cencier A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.B., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush P., Kapen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasmo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Maleshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclele J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard D., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; Pubmed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muny D.M., Nelson C.R.,
RA Paclele J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svitskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.B., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; Pubmed=12537573;
RA Kaminler J.S., Bergman C.M., Krommiller B., Carlson J., Svitskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.,
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; Pubmed=12537572;
RA Miya S., Crosby M.A., Mungall C.J., Matthee B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminler J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tudy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Beltenkourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
[5]
RP SEQUENCE FROM N.A.
RX Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [6]
RP SEQUENCE FROM N.A.
RX Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB003513; AAF49008.1; -
DR Intact; QYVMD6; -
DR FlyBase: FBgn0031060; CG14231.
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase_M22.

DR InterPro; IPR009180; Pept M22_Os1a1g1.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Os1a1g1_ptide; 1.
 DR PRINTS; PR00789; OS1ALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 SQ SEQUENCE 409 AA; 45328 MW; 979766E7D15538 CRC64;

Query Match 26.8%; Score 332.5; DB 2; Length 409;
 Best Local Similarity 35.9%; Pred. No. 6.9e-21;
 Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

DR 2 LALVQGVSDFLLEKSLDIAFGDMLDKVARRLSLIKHECSTSGSKAIEHLAK-QGRNF 60
 DB LVVANGGRLLTLGCTDIDDAFGAEPDCKGRRLHIPEFLYLMNGRAIEHAQIASDPL 226
 QY 61 HFDIKPPLHAAKCDPFGTQVHTDKTIIMKEKEEGIEKQQLSSADIAATVQHTMAC 120
 DB 227 AVEFPLPLAQRNCFNFFAGIKKNSFRAIRARERPRPDGVISNYGDFCAGLLRSYSR 286
 QY 121 HLVKRTIRALFC-KQRDLIPQNNAVLVASGVASNYIRALIELTNATQCTLLCPP 178
 DB 287 HLMKRTQRALEYCLPHRQLFGDTPPTLVMSGVANNDAIYANIEHLAAQYGCSEFRPSK 346
 QY 179 RLCTDNGIMIAMNGIERLRAGLGIHLIDIE-GIRYEPKCPGLGVDSKEVGEA 228
 DB 347 RYCSDNQVMIAHMGVEOL-----LQDKKASTRIDYD--SIDLOGSAGFA 388

RESULT 13
 022145 PRELIMINARY; PRT; 480 AA.

AC 022145; 08VWL2;
 DT 01-JUN-1998 (TrEMBLrel. 05, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease GCP1).
 GN Name=At2g45270; Synonyms=GCP1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NC NCBI_TaxID=3702;

RA [1]
 RP SEQUENCE FROM N.A.
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,
 RA Venter J.C.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Town C.D., Kaul S.;
 RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hausseuhl K., Adamaka I.;
 RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RA Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang B., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,

RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA EMBL; AC002387; AAB82636.2; -
 DR EMBL; AY024338; AAK00530.1; -
 DR EMBL; AY063864; AAL36220.1; -
 DR EMBL; AY117283; AAM51358.1; -
 DR PIR; E84888; E84888.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR00905; Peptidase_M22.
 DR InterPro; IPR009180; Pept M22_Os1a1g1.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Os1a1g1_ptide; 1.
 DR PRINTS; PR00789; OS1ALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 480 AA; 52995 MW; 20DD6A86ACC1FPAD CRC64;

Query Match 26.1%; Score 324; DB 2; Length 480;
 Best Local Similarity 34.1%; Pred. No. 4.7e-20;
 Matches 86; Conservative 34; Mismatches 86; Indels 46; Gaps 6;

DR 1 LALVQGVSDFLLEKSLDIAFGDMLDKVARRLSLIKHECSTSGSKAIEHLAKQGRNF 60
 DB 224 LVLVAKHGQYQLGTYDDAIGEAFDKRAKVLGLDMH-----RSGGPAVEELALGDAK 278
 QY 61 HFDIKPPLHAAKCDPFGTQVHTDKTIIMKEKEEGIEKQQLSSA-----ADIAA 112
 DB 279 SVKFAVPMKHYKDCNFSYAGLTKQVRLAIEAKE-----IDAKCPVSATVEDRRNRADIAA 334
 QY 113 TVQHTMACLVKRTIRALFCQQRDLIPQNNAVLVASGVASNYIRALIELTNATQCT 172
 DB 335 SFQRAVALHLEKRCRAIDMALE--LEPSIKHMVTSIGVASNKVYRLRLNIVENKNTK 391
 QY 173 LDCPPRLCTDNGIMIAMNGIERLRAGLGIHLIDIEGIRYE-----PKCP 216
 DB 392 LVCPFPRLCTDNGVAVMTGLHFRVVG-----RIDPPPAIEPBDYVDIARPWP 441

QY 217 LGVDISKEVGEA 228
 DB 442 LGBEYAKGRSEA 453

RESULT 14
 0677H2 PRELIMINARY; PRT; 255 AA.

AC 0677H2;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE O-sialoglycoprotein endopeptidase (Fragment).
 GN Hyacinthus orientalis.
 OS Hyacinthus orientalis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Hyacinthaceae;
 OC Hyacinthus.
 NC NCBI_TaxID=82025;
 RA [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Floral meristem 5-10 days when regenerated in vitro;
 RA Fan J.H., Ma Y., Zhang X.S.;
 RT mRNA, expressed during regeneration of floral bud."
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY389554; AAT08652.1; -
 DR InterPro; IPR00905; Peptidase_M22.
 DR PRINTS; PR00789; OS1ALOPTASE.

DR Prodom; PD002367; Peptidase_M22; 1.
 FT NON TER 1
 FT NON TER 255
 SQ SEQUENCE 255 AA; 28516 MW; 529D22666C284106 CRC64;

Query Match 24.2%; Score 299.5; DB 2; Length 255;
 Best Local Similarity 37.4%; Pred. No. 3.3e-18;
 Matches 79; Conservative 26; Mismatches 79; Indels 27; Gaps 5;

QY 14 LGKSLDIARGMLDVKARLSLKHPECSMTSGKAIITHLAKQGRFHFDIIRPLHAKN 73
 DB 1 LGTIDIDIGAYDVTARMLGL-----DMGKGGGALBELALEGANSVKFKIPIRQHKD 55
 QY 74 CDPSFTGLQHTYDKIMKEKEKEGKQILSSA-----ADIAATVQHTMACHLVKRT 126
 DB 56 CNFSYAGIK--TQRLAESRNICIDKSPISLATCHDRMRADIAASQVAVLHLEDRG 113
 QY 127 HRAILFCRKQRLDLPONNAVLSGGSVSNFYIRRALILTNATQCTLLCPPRLCTDNGI 186
 DB 114 ARAIEMALIEIPTVQ---CLVVGGSVANSKRYRSRLNHLVVTGLRLVCPSPSLCTDNGV 170
 QY 187 MIANNGIERLARGILHDIEGIRYEPKCPL 217
 DB 171 MVAMTGIENFLVG-----RYDPPPV 191

RESULT 15

Q92LH8 PRELIMINARY; PRT; 360 AA.
 ID Q92LH8
 AC Q92LH8
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).
 GN ORFNames=SMC03230;
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 NC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetle D., Puehler A., Purnelle B., Rameberger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 Sinorhizobium meliloti strain 1021.",
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591792; CAC47657.1; -.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009095; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22_1.
 DR PIRSF; PIRSF004537; Osialglc_ptids; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KM Complete proteome; Hydrolase.
 SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EA86 CRC64;

Query Match 21.2%; Score 262.5; DB 2; Length 360;
 Best Local Similarity 32.8%; Pred. No. 9e-15;
 Matches 78; Conservative 32; Mismatches 85; Indels 43; Gaps 7;

QY 2 LALVQGVSDPFLILKSLDIARGMLDVKARLSLKHPECSMTSGKAIITHLAKQGRFH 61
 DB 145 LILVKGVSERYRWGTTIDDLGELAPDKTKLGL-PYP-----GGPAVERAQAQNAER 197

QY 62 FDIKPELHAKKCDPSFTGLQ-----HYTDKIMKEKEKEGKQILSSADI 110
 DB 198 FPPPRPLVDADALDFSGLKTRAVRQAQSLGPTVTDI-----ADV 239
 QY 111 AATVQHTMACHLVKRTHRAILFCRKQRLDLPONNAVLSGGSVSNFYIRRALILTNATQ 170
 DB 240 CASFGRAISRTLDRVRGRLKFRF--ADFASVDQPALVAVGVAANQTLRRTLQSLCDEHG 298
 QY 171 CTLLCPPRLCTDNGIMIANNGIERLARGILHDIEGIRYEPKC--PLGVDISKVVG 226
 DB 299 FFFIAPPIQLCTDNAMAMIAAGIERLAAGL---PADGLDAAPRSRWPIDSEKALIG 352

Search completed: February 16, 2005, 13:08:41
 Job time : 41.5261 secs